

QY 122 SADVALVELEAPVETNYILPVLDPSPVIFETGMCWTVGMC--SPSEEDILPEPRILQ 179
 DB 132 QGDALILQSRPITFSRIRPILPLPAMNSFRGLHCTVGMCHVAPSVSLTPKP--LQ 189
 QY 180 KLVPIIDITPKMLYSKDFEFGYQPKTKNDMLCAGFEFGKKDCKGSGGPLYVGLVGO 239
 DB 190 QLEPPLISRFETCKLYNIDAK--PEEPHFVQEDVACAGYEGGKDCQSGGPLYSCFVGG 248
 QY 240 SWIAGVISWEGCAGARONRPGVYIRVTAHNMH-----RIIPKIQ 280
 DB 249 LMTLTGIVSWGACGARNRPGVYITLASSVSIQSKVTELDQRPVHPOTQ 297

RESULT 2
 S56160
 mast cell tryptase precursor - Mongolian jird
 C:Species: Meriones unguiculatus (Mongolian jird)
 C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 Accession: S56160
 EMBL: J. 309, 921-926, 1995
 A:Reference number: S56160; MUID:9556971
 A:Accession: S56160
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-270 <MUR>
 A:Cross-references: EMBL:D31789; NID:G517122; PIDN:BA06598.1; PID:G517123
 C:Superfamily: trypsin; trypsin homology
 F:26-262/Domain: trypsin homology <TRY>

Query Match 37.8%; Score 594; DB 2; Length 270;
 Best Local Similarity 44.0%; Pred. No. 7,1e-45;
 Matches 122; Conservative 48; Mismatches 87; Indels 20; Gaps 7;
 QY 6 AVPLLLIFGSGORAKATACGRPRLNRMVGGDTGCEMPVOVSIORNGS--HECGG 62
 DB 8 ALPLSLM-----HNSPLCOEWG-----IVGGEPAPGNKMPWOVSLNANETVMHFCGG 56
 QY 63 SLIAQWVLTAAHCPFRNT--SETSILYOVLLGAROLYQPGPHAMVARYOVESNPLYOGFAS 121
 DB 57 SLIHQWVLTAAHCVGPRLADPNKRVQLRKQYLY--HDLAVNSITTHPTFYATQ 113
 QY 122 SADVALVELEAPVETNYILPVLDPSPVIFETGMCWTVGMCSPSEEDILPEPRILQ 181
 DB 114 GADIALLELKMPVNISSHVHVSILPPASETPPGSLCWTGMDNDVSLPPFPPLKEY 173
 QY 182 AVPIIDITPKMLYSKDFEFGYQPKTKNDMLCAGFEFGKKDCKGSGGPLYVGLVGO 241
 DB 174 QVPVENQICDLKHKYGTGDNHIVADMLCAG--NEG--HDSQSGSGGPLYVCKVNGTW 231
 QY 242 LQAGVISWEGCAGARONRPGVYIRVTAHNMHRIIPK 278
 DB 232 LQAGVISWEGCAGALNRPGRVYITVYLLDMHRYVVK 268

RESULT 3
 B35863
 tryptase (EC 3.4.21.59) II precursor - human
 N:Alternate names: tryptase beta
 C:Species: Homo sapiens (man)
 C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000
 C:Accession: B35863; A37193; I59473
 R:Vanderlisse, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
 A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr
 A:Reference number: A35863; MUID:90251647
 A:Accession: B35863
 A:Molecule type: mRNA
 A:Residues: 1-275 <VAN>
 A:Cross-references: GB:M33492; NID:G339982; PIDN:AAA36779.1; PID:G339983

A>Note: residues 2-275 are derived from mRNA; residue one was inferred from the genom
 R:Miller, U.S.; Moxley, G.; Schwartz, L.B.
 J. Clin. Invest. 86, 864-870, 1990
 A:Title: Cloning and characterization of a second complementary DNA for human tryptas
 A:Reference number: A37193; MUID:90369005
 A:Accession: A37193
 A:Molecule type: mRNA
 A:Residues: 1-275 <ML>
 A:Cross-references: GB:M37488; NID:G179583; PIDN:AAA51843.1; PID:G179584
 R:Blom, T.; Hellman, L.
 Scand. J. Immunol. 37, 203-208, 1993
 A:Title: Characterization of a tryptase mRNA expressed in the human basophil cell lin
 A:Reference number: I59473; MUID:93166209
 A:Accession: I59473
 A:Status: preliminary; translated from GB/EMBL/DDBY
 A:Molecule type: mRNA
 A:Residues: 1-275 <RPS>
 A:Cross-references: GB:S5551; NID:G265666; PIDN:AA013876.1; PID:G4261576
 A:Experimental source: basophil cell line KU812
 C:Genetics:
 A:Gene: GDB:TPS1
 A:Cross-references: GDB:125890; OMIM:191080
 A:Map position: 16pter-16qter
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; serine proteinase; zymogen
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-30/Domain: activation peptide #status predicted <ACT>
 F:31-275/Product: tryptase I #status predicted <MAT>
 F:31-267/Domain: trypsin homology <TRY>
 F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 37.5%; Score 588.5; DB 2; Length 275;
 Best Local Similarity 43.8%; Pred. No. 2.2e-44;
 Matches 120; Conservative 46; Mismatches 99; Indels 9; Gaps 4;
 QY 9 LLLILFSGORAKATACGRPRLNRMVGGDTGCEMPVOVSIORNGS--SHECGGSLI 65
 DB 5 LLALPLARAVAAAPRAGALRVGLVGGELAPRSKTRQVSLRKHGTYMHHFCGSLI 64
 QY 66 AEDWVLTAAHCF--RNTSESLYOVLLGAROLYQPGPHAMVARYOVESNPLYOGFAS 124
 DB 65 HPQWVLTAAHCVGPRLADPNKRVQLRKQYLY--QDLPVSRITVHPQFYTAQIGAD 121
 QY 125 VALVELEAPVETNYILPVLDPSPVIFETGMCWTVGMCSPSEEDILPEPRILQ 184
 DB 122 IALLELEPRKVSNNHTYTLPPASETPPGMPCWTVGMDVNDERLPPPLKQVKKP 181
 QY 185 IDTPKCNLLYSKDFEFGYQPKTKNDMLCAGFEFGKKDCKGSGGPLYVGLVGO 244
 DB 182 IMENHICDAKHYHGAVTGDVIRVDDMLCAG--NTRRSCGSGGPLYVCKVNGTW 239
 QY 245 GVVISWEGCAGARONRPGVYIRVTAHNMHRIIPK 278
 DB 240 GVVISWEGCAGARONRPGVYIRVTAHNMHRIIPK 273

RESULT 4
 A35863
 tryptase (EC 3.4.21.59) I precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 21-Jul-2000
 C:Accession: A35863; D35863; A60939; A39326
 R:Vanderlisse, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
 A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine
 A:Reference number: A35863; MUID:90251647
 A:Accession: A35863
 A:Molecule type: DNA
 A:Residues: 1-275 <VAN>
 A:Cross-references: GB:M33494; NID:G3927804; PIDN:AAC63172.1; PID:G339977
 A:Accession: D35863
 A:Molecule type: mRNA

Query Match 37.0%; Score 581; DB 2; Length 274;
 Best Local Similarity 43.8%; Pred. No. 1e-43;
 Matches 124; Conservative 38; Mismatches 93; Indels 28; Gaps 5;

9 LLLLLCFSQAKAKATACGRPRMLNRWVGODTOEGEMPOVSIQ---RNGSHFCGSLI 65
 Db 5 LLLALLPVASAKAYAPAPVQALQAGIVGGEAPRSKMPQVSLRVDRTVMHFCGSLI 64
 QY AEGWVLTAAHCFRNTSETSLYQVLLG-----ARQLVQCPHANY---ARVQVESNPL 115
 Db 65 HPQWVLTAAHCFRNTSETSLYQVLLG-----LGRVDVLDLALRYNSGTHLYQDQLPVSRIWHQ 111
 QY 116 YQTAASADVALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGPSEEDLLPER 175
 Db 112 FYIQTGADIALLELEFPVNSRVHTVMLPPASEFPFPPMPCWVGMDVNDDEPLRP 171
 QY 176 RILOKLVPIIDPKKLLYSKDFEYQYKTIKNDMLCAGFEGBKKDCKGSGGGLVC 235
 Db 172 FPLQVAVPPIEMNHICAKYHLGAYTGDDVRIIRDMCLAG--NSQRDCKGSGGGLVC 229
 QY 236 LVQGSMLQAGVISMGEGCARQNRPGVYIRYTAHNNHRIIRPK 278
 Db 230 KVNCTWLAGVWSMGEGCARQNRPGVYIRYTAHNNHRIIRPK 272

RESULT 7
 A47246
 tryptase (EC 3.4.21.59) 2 - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A47246
 R:McNell, H.P.; Reynolds, D.S.; Schiller, V.; Ghildyal, N.; Gurely, D.S.; Austen, K.F.;
 Proc. Natl. Acad. Sci. U.S.A. 89, 11174-11178, 1992
 A:Title: Isolation, characterization, and transcription of the gene encoding mouse mast
 A:Reference number: A47246; MUID:93087489
 A:Accession: A47246
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-273 <MCN>
 A:Cross-references: GB:I00653; NID:9200518; PIDN:AAA39982.1; PID:9200519
 A:Note: Sequence extracted from NCBI backbone (NCBI:119745, NCBI:119746)
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; trypsin proteinase
 F:29-265/Domain: trypsin homology <TRY>

Query Match 36.8%; Score 578.5; DB 2; Length 273;
 Best Local Similarity 45.5%; Pred. No. 1.7e-43;
 Matches 125; Conservative 41; Mismatches 96; Indels 13; Gaps 7;

9 LLLLLCFSQAKAKATACGRPRMLNRWVGODTOEGEMPOVSIQNGS---HFCGSL 64
 Db 5 LLLTLPPLSLVHAAPG---PAMTREGIVGGEAHGNKMPQVSLRANDTVMHFCGSL 61
 QY 65 IAEQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLYQGTASSA 123
 Db 62 IHPQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLYQGTASSA 118
 QY 124 DYALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGPSEEDLLPERILQKLV 183
 Db 119 DIALKLTNPVNIISYVHVPPLPPASEFPFPPMPCWVGMDVNDDEPLRP 178
 QY 184 PIIDPKKLLYSKDFEYQYKTIKNDMLCAGFEGBKKDCKGSGGGLVC 243
 Db 179 PIIEHLDLCKYHKLITGDVNIHYRDMCLAG--HDSQGGSGGLVC 236
 QY 244 AGVSWGEGCARQNRPGVYIRYTAHNNHRIIRPK 278
 Db 237 AGVSWGEGCARQNRPGVYIRYTAHNNHRIIRPK 271

RESULT 8
 JC4171

tryptase (EC 3.4.21.59) precursor - rat
 N:Alternate names: mast cell tryptase
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 27-Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000
 C:Accession: JC4171
 R:Ido, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.
 J. Biochem. 118, 210-215, 1995
 A:Title: cDNA sequencing and expression of rat mast cell tryptase.
 A:Reference number: JC4171; MUID:96015171
 A:Accession: JC4171
 A:Molecule type: mRNA
 A:Residues: 1-274 <IDB>
 A:Cross-references: DDBJ:D38455; NID:9556555; PIDN:BA07486.1; PID:9556556
 C:Comment: This enzyme is basically specific for a connective tissue mast cell, it is
 elase inhibitors.
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; mast cell; serine proteinase; zymogen
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-29/Domain: activation peptide #status predicted <ACT>
 F:30-274/Product: mast cell tryptase #status predicted <MAT>
 F:30-266/Domain: trypsin homology <TRY>
 F:73,120,223/Active site: His, Asp, Ser #status predicted
 F:131/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 36.0%; Score 565.5; DB 2; Length 274;
 Best Local Similarity 43.7%; Pred. No. 2.4e-42;
 Matches 122; Conservative 41; Mismatches 97; Indels 19; Gaps 7;

9 LLLLLCFSQAKAKATACGRPRMLNRWVGODTOEGEMPOVSIQNGS---HFCGSLI 65
 Db 4 LLLTLPPLSLVHAAPG---PAMTREGIVGGEAHGNKMPQVSLRANDTVMHFCGSLI 63
 QY 66 IAEQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLYQGT 120
 Db 64 IHPQWVLTAAHCFRNTSETSLYQVLLGARQLVQCPHANYRQVESNPLYQGT 116
 QY 121 SSADVALVELAPVFTNYILPVCPLDPSVIFETGMNCWVGSGPSEEDLLPERILQ 179
 Db 117 DGADIALLELEFPVNSRVHTVMLPPASEFPFPPMPCWVGMDVNDDEPLRP 175
 QY 180 KLVAVPIIDPKKLLYSKDFEYQYKTIKNDMLCAGFEGBKKDCKGSGGGLVC 239
 Db 176 QVQVPIEENSICDKRYHTGTYGDDVPIYQGMCLAG--NTRSDSCGGSGGLVC 233
 QY 240 SMLQAGVISMGEGCARQNRPGVYIRYTAHNNHRIIRPK 278
 Db 234 TWLQAGVWSMGEGCARQNRPGVYIRYTAHNNHRIIRPK 272

RESULT 9
 A38654
 mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Feb-1992 #sequence_revision 17-Feb-1994 #text_change 22-Jun-1999
 C:Accession: A38654; B38654; D35646; I59478
 R:Reynolds, D.S.; Gurely, D.S.; Austen, K.F.; Serafin, W.E.
 J. Biol. Chem. 266, 3847-3853, 1991
 A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by
 A:Reference number: A38654; MUID:91139682
 A:Accession: A38654
 A:Molecule type: DNA
 A:Residues: 1-276 <RE2>
 A:Cross-references: GB:M57625; NID:9200506; PIDN:AAA39987.1; PID:9200507
 A:Note: the authors translated the codon CGC for residue 24 as Ala, GAG for residue 3
 s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly
 A:Accession: B38654
 A:Molecule type: mRNA
 A:Residues: 1-276 <RE2>
 A:Cross-references: GB:M57626; NID:9200508; PIDN:AAA39988.1; PID:9200509
 R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.
 Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990
 A:Title: Different mouse mast cell populations express various combinations of at least

Wed Aug 14 08:39:23 2002

us-10-041-006-7.rpr

Page 6

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:52:53 ; Search time 13.11 Seconds
(without alignments)
540.307 Million cell updates/sec

Title: US-10-041-006-7
Perfect score: 1571
Sequence: 1 MRRAVAVPLLLLCFSGSQA.....WIRRIIPKIQFQPARLGQK 290

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 231628
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 1197 | 76.2 | 304 | 4 | US-09-088-651-2 |
| 2 | 602.5 | 38.4 | 299 | 4 | US-08-944-483-66 |
| 3 | 594 | 37.8 | 270 | 2 | US-08-978-404B-8 |
| 4 | 590.5 | 37.6 | 273 | 2 | US-08-978-404B-6 |
| 5 | 588.5 | 37.5 | 274 | 2 | US-09-016-366A-21 |
| 6 | 588.5 | 37.5 | 274 | 2 | US-08-978-404B-16 |
| 7 | 587.5 | 37.4 | 273 | 2 | US-09-016-366A-19 |
| 8 | 587.5 | 37.4 | 273 | 2 | US-08-978-404B-14 |
| 9 | 578.5 | 36.8 | 267 | 2 | US-09-016-366A-23 |
| 10 | 578.5 | 36.8 | 267 | 2 | US-08-978-404B-18 |
| 11 | 578.5 | 36.8 | 273 | 2 | US-08-978-404B-3 |
| 12 | 574 | 36.5 | 314 | 4 | US-09-008-271A-3 |
| 13 | 573.5 | 36.5 | 249 | 4 | US-09-079-970A-5 |
| 14 | 572.5 | 36.4 | 275 | 2 | US-09-016-366A-17 |
| 15 | 572.5 | 36.4 | 275 | 2 | US-08-978-404B-12 |
| 16 | 568.5 | 36.2 | 245 | 4 | US-09-079-970A-6 |
| 17 | 565.5 | 36.0 | 274 | 2 | US-08-978-404B-5 |
| 18 | 564 | 35.9 | 276 | 2 | US-09-016-366A-15 |
| 19 | 564 | 35.9 | 276 | 2 | US-08-978-404B-21 |
| 20 | 560.5 | 35.7 | 245 | 4 | US-08-944-483-69 |
| 21 | 527.5 | 33.6 | 248 | 4 | US-08-944-483-63 |
| 22 | 523.5 | 33.3 | 416 | 2 | US-09-000-846-2 |
| 23 | 517 | 32.9 | 638 | 2 | US-08-681-151-3 |
| 24 | 505 | 32.1 | 791 | 1 | US-08-643-219-1 |
| 25 | 505 | 32.1 | 791 | 3 | US-08-851-350-1 |
| 26 | 504 | 32.1 | 256 | 2 | US-09-027-337-3 |
| 27 | 503 | 32.0 | 814 | 1 | US-08-750-711-1 |

| | | | | | | |
|----|-------|------|-----|---|-------------------|--------------------|
| 28 | 502 | 32.0 | 790 | 1 | US-08-469-486-54 | Sequence 54, Appl |
| 29 | 502 | 32.0 | 790 | 2 | US-08-469-658-54 | Sequence 54, Appl |
| 30 | 502 | 32.0 | 791 | 2 | US-09-131-995-1 | Sequence 1, Appl1 |
| 31 | 502 | 32.0 | 791 | 2 | US-08-832-087B-1 | Sequence 1, Appl1 |
| 32 | 502 | 32.0 | 791 | 4 | US-09-132-154-1 | Sequence 1, Appl1 |
| 33 | 502 | 32.0 | 810 | 1 | US-07-854-603-2 | Sequence 2, Appl1 |
| 34 | 502 | 32.0 | 810 | 1 | US-08-147-000B-29 | Sequence 29, Appl1 |
| 35 | 502 | 32.0 | 810 | 4 | US-09-086-514-1 | Sequence 1, Appl1 |
| 36 | 502 | 32.0 | 810 | 6 | 5200340-8 | Patent No. 5200340 |
| 37 | 499 | 31.8 | 255 | 4 | US-08-944-483-67 | Sequence 67, Appl |
| 38 | 499 | 31.8 | 546 | 6 | 5200340-6 | Patent No. 5200340 |
| 39 | 493 | 31.4 | 269 | 2 | US-08-978-404B-10 | Sequence 10, Appl |
| 40 | 483.5 | 30.8 | 230 | 1 | US-08-456-840-47 | Sequence 47, Appl |
| 41 | 483.5 | 30.8 | 230 | 1 | US-08-266-407A-47 | Sequence 47, Appl |
| 42 | 483.5 | 30.8 | 230 | 2 | US-08-892-544-47 | Sequence 47, Appl |
| 43 | 483.5 | 30.8 | 230 | 2 | US-08-766-983-12 | Sequence 12, Appl |
| 44 | 483.5 | 30.8 | 230 | 4 | US-08-944-483-53 | Sequence 53, Appl |
| 45 | 483.5 | 30.8 | 230 | 4 | US-09-296-219-12 | Sequence 12, Appl |

ALIGNMENTS

RESULT 1
US-09-088-651-2
; Sequence 2, Application US/09088651
; Patent No. 6165771
; GENERAL INFORMATION:
; APPLICANT: BURGESS, NICOLA A.
; APPLICANT: CLINKENBEARD, HELEN E.
; APPLICANT: SOUTHAN, CHRISTOPHER D.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RAINER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/088,651
; FILING DATE: JUNE 1, 1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB9712088.5
; FILING DATE: 10-JUNE-1997
; APPLICATION NUMBER: EP 97308295.1
; FILING DATE: 17-OCT-1997
; APPLICATION NUMBER: GB 9803650.2
; FILING DATE: 20-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F.
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH30358
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-088-651-2

Query Match 76.2%; Score 1197; DB 4; Length 304;
 Best Local Similarity 82.1%; Pred. No. 1.5e-124;
 Matches 234; Conservative 7; Mismatches 36; Indels 8; Gaps 3;

QY 1 MRRPAVPLLLLCGSGRANAKATACGRPRMLNRVYGODTQEGEMPQVSIORNGSHFC 60
 DB 1 MRRPAVPLLLLCGSGRANAKATACGRPRMLNRVYGODTQEGEMPQVSIORNGSHFC 60
 QY 61 GGLAEQWVLTAAHCENFTSETSLXYLLARQLVQGPAMARARQVESNPLYOGTA 120
 DB 61 GGLAEQWVLTAAHCENFTSETSLXYLLARQLVQGPAMARARQVESNPLYOGTA 120
 QY 121 SSADVALVELAPVFTYIIPVCLPDPSTVLEFGMNCWVGWSPSEDLPEPRILQK 180
 DB 121 SSADVALVELAPVFTYIIPVCLPDPSTVLEFGMNCWVGWSPSEDLPEPRILQK 180
 QY 181 LAVPIIDPKCNLYSKDTEFGYQPKTIKNDMLCAGEFEKKKACGDSGPLYCLVQGS 240
 DB 181 LAVPIIDPKCNLYSKDTEFGYQPKTIKNDMLCAGEFEKKKACGDSGPLYCLVQGS 240
 241 WLQAGV-1SMGEG-----CARONRPGYIRYTAHNNHRIIPKL 279
 239 FQPIGPGGKGLGPPVWMLQTEPPSVLIRVNGOONMTHRIIPKL 283

RESULT 2
 US-08-944-483-66
 ; Sequence 66, Application US/08944483
 ; Patent No. 6232456

GENERAL INFORMATION:
 APPLICANT: COHEN, MAURICE
 APPLICANT: COLPITTS, TRACEY L.
 APPLICANT: FRIEDMAN, PAULA N.
 APPLICANT: GRANADOS, EDWARD N.
 APPLICANT: KLAAS, MICHAEL R.
 APPLICANT: RUSSELL, JOHN C.
 APPLICANT: STEWART, KENT D.
 APPLICANT: STROUPE, STEVEN D.
 TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
 NUMBER OF SEQUENCES: 76
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Abbott Laboratories
 STREET: 100 Abbott Park Road
 CITY: Abbott Park
 STATE: IL
 COUNTRY: USA
 ZIP: 60064-3500

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/944,483
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
 NAME: Becker, Cheryl L.
 REGISTRATION NUMBER: 35,441
 REFERENCE/DOCKET NUMBER: 6183.US.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 847/935-1729
 TELEFAX: 847/938-2623
 TELEX:
 INFORMATION FOR SEQ ID NO: 66:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 299 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: No. 62324566
 US-08-944-483-66

Query Match 38.4%; Score 602.5; DB 4; Length 299;
 Best Local Similarity 43.8%; Pred. No. 1.2e-58;
 Matches 112; Conservative 44; Mismatches 87; Indels 13; Gaps 4;

QY 35 MVGQTOEGEMWQVSIORNGSHFCGSLIAQWVLTAAHCENFTSETSLXYLLARQ 94
 DB 1 ITGSSNVAQWQWQVSTIEGVHVCGLSVQWVLSAHCENFTSETSLXYLLARQ 60
 QY 95 LVQGPAMARARQVESNPLYOGTASADVALVELAPVFTYIIPVCLPDPSTVLEF 154
 DB 61 LQSYSDAKVSTLKDILPHPSYLOEGSGDIALQLSRPTFSRYIRPCLPAANASFPN 120
 QY 155 GNMCMWTGNC--SPSEDLPEPRILQKLAVIDPKCNLYSKDTEFGYQPKTIKNDM 212
 DB 121 GLHCTVIGWHAAPSLSLTPKP--LQGLEVPLISRETNCCLYNIDAK--PEEPHFVQEDM 177
 QY 213 ICAGEFEKKKACGDSGPLYCLVQGSWLAQVLSGEGCARONRPGYIRYTAHNNWI 272
 DB 178 VCAGYVEGKDKACQDSGGLSCPVVGLMYLIGVSMGACARNRPGYITLASSYASHI 237
 QY 273 H-----RIIPKLQ 280
 DB 238 QSKVTELQPRVVPQTQ 253

RESULT 3
 US-08-978-404B-8
 ; Sequence 8, Application US/08978404B
 ; Patent No. 5968782

GENERAL INFORMATION:
 APPLICANT: Stevens, Richard L.
 TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
 TITLE OF INVENTION: FIBRINOGEN
 NUMBER OF SEQUENCES: 74
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wolf, Greenfield & Sachs, P.C.
 STREET: 600 Atlantic Avenue
 CITY: Boston
 STATE: MA
 COUNTRY: U.S.A.
 ZIP: 02210-2211

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,404B
 FILING DATE: 25-NOV-97
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/032,354
 FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:
 NAME: Plummer, Elizabeth R.
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B0801/7090
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 TELEX:

INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 270 amino acids
 TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: No. 5968782e
US-08-978-404B-8

Query Match 37.8%; Score 594; DB 2; Length 270;
Best Local Similarity 44.0%; Pred. No. 9.1e-58;
Matches 122; Conservative 48; Mismatches 87; Indels 20; Gaps 7;

6 AVPLLLLCGSGRAKATACGRRLNRMVGGODTOEGEMPWQVSIQRNGS---HFCGG 62
8 ALPLFSLM-----HRSPLCQEMG-----IVGQEARGNKMPQVSLRANETWMHFCGG 56
63 SLIAEOWVLAHACFRNT-SETSLYOVLGAROLVOPGHAMTARYKQVESNPVLCGRASSND 121
57 SLIHQWVLAHACVGPPLIADPNKRVOLRKQYLXY---HDHLLAVSRITTHPTFYATON 113
122 SADVALVLEAPVPTNTIIPVCLPDPSPVFEFGMNCWTGWCSPSEEDLLPEPRILOKL 181
114 GADIALLELKNPVNISSHVPSLPPASETFPSCGLCWYTGKGNINDVSLPPFPPLKEV 173
182 AVPIIDFKCNILYSKDTERGYPKTIKNDMLCAGFEEGKADCKDGGSGPLVCLVGSQSW 241
174 QVPVVENLCLDKYHKGYTSDNTHIVRDMCLAG-NEG-HDSCGDSGSGPLVCKVNGTW 231
242 LGAGVTSWEGECARONRGVYIRYTAHNMTHRIIPK 278
232 LGAGVTSWEGECALPNRPGIYTRVYLYLDMTHRYVPK 268

RESULT 4
US-08-978-404B-6
Sequence 6, Application US/08978404B

PATENT No. 5968782
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
TITLE OF INVENTION: FIBRINOGEN
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5968782e

US-08-978-404B-6

Query Match 37.6%; Score 590.5; DB 2; Length 273;
Best Local Similarity 45.6%; Pred. No. 2.3e-57;
Matches 125; Conservative 43; Mismatches 95; Indels 11; Gaps 6;

9 LLLLCGSGRAKATACGRRLNRMVGGODTOEGEMPWQVSIQRNGS---HFCGSLI 65
5 LLLTPLLSLVHAAPSLAMP--EGIVGQEARGNKMPQVSLRANETWMHFCGSLI 62
66 AEQWVLAHACF-RNTSETSLYOVLGAROLVOPGHAMTARYKQVESNPVLCGRASSND 124
63 HPQWVLAHACVGNKADPNKRLVQLRKQYLXY---HDHLLVTSQIISHPDEFIAODGAD 119
125 VALVLEAPVPTNTIIPVCLPDPSPVFEFGMNCWTGWCSPSEEDLLPEPRILOKLAVP 184
120 IALLKTNPVNITSNVHTVSLPPASETFPSCGLCWYTGKGNINDVSLPPFPPLKEVQVP 179
185 IIDPFKCNILYSKDTERGYPKTIKNDMLCAGFEEGKADCKDGGSGPLVCLVGSQSWLDA 244
180 IVENRLCDLKYHKGLNTGDNVHIVRDMCLAG-NEG-HDSCGDSGSGPLVCKVEDTWLDA 237
245 GVTWEGECARONRGVYIRYTAHNMTHRIIPK 278
238 GVTWEGECALPNRPGIYTRVYLYLDMTHRYVPK 271

RESULT 5
US-09-016-366A-21
Sequence 21, Application US/09016366A

PATENT No. 5955431
GENERAL INFORMATION:
APPLICANT: Huang, Chifu
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
TITLE OF INVENTION: INHIBITORS
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,366A
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-016-366A-21

37.58; Score 588.5; DB 2; Length 274;

| | | | | |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match | 37.48; | Score 587.5; | DB 2; | Length 273; |
| Best Local Similarity | 43.88; | Pred. No. 4.9e-57; | | |
| Matches 120; | Conservative 46; | Mismatches 99; | Indels 9; | Gaps 4; |



| | | |
|----|---|--|
| XX | AA73388 | |
| ID | AAV73388 standard; Protein; 290 AA. | |
| XX | | |
| AC | AAV73388; | |
| XX | | |
| DT | 24-FEB-2000 (first entry) | |
| XX | | |
| DE | HTRM clone 3376404 protein sequence. | |
| XX | | |
| KW | HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS | |
| KW | arteriosclerosis; cirrhosis; cancer; leukaemia; diabetes mellitus; | |
| KW | Addison's disease; multiple sclerosis; rheumatoid arthritis; infection; | |
| XX | trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment. | |
| OS | Homo sapiens. | |
| XX | | |
| PN | WO957144-A2. | |
| XX | | |
| PD | 11-NOV-1999. | |
| XX | | |
| PF | 04-MAY-1999; 99WO-US09935. | |
| XX | | |
| PR | 05-MAY-1998; 98US-0084254. | |
| PR | 07-AUG-1998; 98US-0095827. | |
| PR | 02-OCT-1998; 98US-0102745. | |
| XX | | |
| PA | (INCY-) INCYTE PHARM INC. | |
| XX | | |
| PI | Hillman JL, Bandman O, Lai P, Yue H, Reddy R, Tang YT; | |
| PI | Gerstlin EH, Patterson C, Baughn MR, Azimzal Y, Lu DM; | |
| XX | | |
| DR | WPI: 2000-052941/04. | |
| XX | N-PSSB; AA252473. | |

XX New peptides useful for diagnosis, prevention and treatment of cancer
 PT and immune disorders -
 XX

PS Claim 1; Page 145-146; 193pp; English.

CC AAV73325-Y73389 are human transcriptional regulator molecule (HTRM)
 CC protein sequences. The HTRM protein and nucleotide sequences are useful
 CC for preventing or treating disorders associated with decreased expression
 CC or activity of HTRM which include cell proliferative disorders such as
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and
 CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists
 CC of the HTRM polypeptides are useful for treating or preventing disorders
 CC associated with increased expression or activity of HTRMs. HTRM
 CC polypeptides, their immunogenic fragments or oligopeptides are useful
 CC for screening libraries of compounds in drug screening techniques.
 CC Polynucleotides encoding HTRM are useful for blocking the transcription
 CC of mRNA and regulating gene function by modulating the activity of
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or
 CC treat disorder associated with decreased HTRM expression. Antibodies
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful
 CC for diagnosing disorders associated with the expression of HTRM,
 CC particularly in assays that detect the expression of HTRM. Nucleotide
 CC sequences encoding HTRM may be useful to generate hybridization probes
 CC useful in mapping the naturally occurring genomic sequence and to detect
 CC differences in gene sequences among normal, carrier and affected
 CC individuals. Using diagnostic assays, cancer can be detected prior to
 CC the appearance of clinical symptoms and thereby progression of cancer can
 CC be prevented by aggressive treatment or preventive measures.

XX Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 21; Length 290;
 Best Local Similarity 100.0%; Pred. No. 1.3e-134;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRRAVPLLLILFCGSRKAKATACGRPMILRMVGGDTGEGMPWOVSIDRNGSHFC 60
 Db 1 MRRRAVPLLLILFCGSRKAKATACGRPMILRMVGGDTGEGMPWOVSIDRNGSHFC 60
 QY 61 GGSLLAEQWVLTAKCFRNTSESLYQVLLGARQVOPGHMAYARVQESNPLXGTA 120
 Db 61 GGSLLAEQWVLTAKCFRNTSESLYQVLLGARQVOPGHMAYARVQESNPLXGTA 120
 QY 121 SSADVALLVELEAVPPTNTYLPVCLPDPSTYIETGNNCWVGTGSGSEEDLPEPTIIOK 180
 Db 121 SSADVALLVELEAVPPTNTYLPVCLPDPSTYIETGNNCWVGTGSGSEEDLPEPTIIOK 180
 QY 121 SSADVALLVELEAVPPTNTYLPVCLPDPSTYIETGNNCWVGTGSGSEEDLPEPTIIOK 180
 Db 121 SSADVALLVELEAVPPTNTYLPVCLPDPSTYIETGNNCWVGTGSGSEEDLPEPTIIOK 180
 QY 241 WLOAGVISMGEGARONRGVYIRYTAHNNHRIITPKILOPAPRLGGCK 290
 Db 241 WLOAGVISMGEGARONRGVYIRYTAHNNHRIITPKILOPAPRLGGCK 290
 QY 241 WLOAGVISMGEGARONRGVYIRYTAHNNHRIITPKILOPAPRLGGCK 290
 Db 241 WLOAGVISMGEGARONRGVYIRYTAHNNHRIITPKILOPAPRLGGCK 290

RESULT 2
 AAU12282
 ID AAU12282 standard; Protein; 290 AA.

XX AAU12282;

XX 24-OCT-2001 (first entry)

XX Human PRO4327 polypeptide sequence.

XX Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;

KW adipocyte; A-peptide; factor VIIA; gene therapy.

XX Homo sapiens.

XX WO200140466-A2.

XX 07-JUN-2001.

PF 01-DEC-2000; 2000WO-US32678.

XX 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 20-MAR-2000; 2000WO-US07377.

PR 21-MAR-2000; 2000WO-US07532.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 10-NOV-2000; 2000WO-US30873.

XX (GETH) GENENTECH INC.

XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart JA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-408281/43.

DR N-PSDB; AAS21354.

XX Claim 12; Fig 222; 813pp; English.

XX AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC or T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,

transgenic or knock out animals and can be used in gene therapy.

Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 22; Length 290;
Best Local Similarity 100.0%; Pred. No. 1,3e-134;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRRPAVPLLLLCFSGSRAKATACGRPMINRMVGGDTQEGEMPWVS:IQRNGSHFC 60
1 mirpaevplllllcfsgsraakatacgrpmlnrmvsgdqtqegewpvs:ldngshfc 60
61 GGSLLAEQWVLTAAHCFRNTSESLVYVLLGAROLVQPPHMYARVROVESNPLYQGT 120
61 ggsllaeqvwltahcfntsestlyvllgarqlvpphmyarvrvsesnplyqgta 120
121 SSADVALVELEAVPFTNYLPVCLDPPSVIFETGNKCVWTGSGSPSEBDLPEPRILQK 180
121 ssadvalveleavpftnylpyclpdpssvifetgnkcvwtgsgspsebdllpeprilqk 180
181 LAVPIIDTPKCNLLYSKDTFEGYQPTIKNDMLCAGFEBGKDACGDSGGLVCLVGS 240
181 lavpidtpkcnnllyskdtfegyqptikndmlcagfeegkdkacgdsqgplvclvgs 240
241 WLQAGVTSNMGEGCARONRPGVYTRVTAHNMWIRHRIIPKLOFORARLGOK 290
241 wlqagvtsnmgegcaronrpgvytrvtahnmwirhriipklofparlgok 290

RESULT 3

AAE03821 standard; Protein; 290 AA.

AAE03821:

08-AUG-2001 (first entry)

Human gene 4 encoded secreted protein HMTH10, SEQ ID NO: 67.

Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
fetal abnormality; developmental abnormality; haematopoietic disorder;
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
inflammation; neurological disorder; Alzheimer's disease; food additive;
pregnancy-related disorder; kidney disorder; gastrointestinal disorder; allergy;
cell culture; chemotaxis; vulnery; binding partner identification;
gene therapy.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..22

Protein /label= Signal_peptide 23..290

WO200136440-A1.

25-MAY-2001.

15-NOV-2000; 2000WO-US31282.

19-NOV-1999; 99US-0166414.

21-JUL-2000; 2000US-0219665.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Komatsoulis GA, Birse CE, Moore PA;
WPI, 2001-343795/36.

N-PSDB: AAD08286.

Isolated nucleic acid molecule encoding a human secreted protein 1s
used in preventing, treating or ameliorating a medical condition -

Claim 11; Page 482-483; 553pp; English.

AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
protein genes, and AAE03818-AAE03870 represent the proteins they encode.
AAE03871-AAE03896 represent human secreted protein fragments or variants.
The secreted proteins and their genes are useful for preventing,
treating or ameliorating medical conditions, e.g., by preventing or gene
therapy. Pathological conditions can be diagnosed by determining the presence of
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
23 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
proliferative disorders, cancer, tumours, foetal and developmental
abnormalities, haematopoietic disorders, diseases of the immune system,
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
allergies, neurological disorders (e.g., Alzheimer's disease,
Parkinson's disease), cognitive disorders, schizophrenia, asthma,
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
cardiovascular disorders, angiogenic disorders, kidney disorders,
gastrointestinal disorders, pregnancy-related disorders, endocrine
disorders, and infections. The proteins can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues, to identify their
cognate ligands or binding partners, and in chemotaxis, and can be used
as a food additive or preservative to modify storage properties.
Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
secreted protein of the invention.

Sequence 290 AA:

Query Match 100.0%; Score 1571; DB 22; Length 290;
Best Local Similarity 100.0%; Pred. No. 1,3e-134;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MRRPAVPLLLLCFSGSRAKATACGRPMINRMVGGDTQEGEMPWVS:IQRNGSHFC 60
1 mirpaevplllllcfsgsraakatacgrpmlnrmvsgdqtqegewpvs:ldngshfc 60
61 GGSLLAEQWVLTAAHCFRNTSESLVYVLLGAROLVQPPHMYARVROVESNPLYQGT 120
61 ggsllaeqvwltahcfntsestlyvllgarqlvpphmyarvrvsesnplyqgta 120
121 SSADVALVELEAVPFTNYLPVCLDPPSVIFETGNKCVWTGSGSPSEBDLPEPRILQK 180
121 ssadvalveleavpftnylpyclpdpssvifetgnkcvwtgsgspsebdllpeprilqk 180
181 LAVPIIDTPKCNLLYSKDTFEGYQPTIKNDMLCAGFEBGKDACGDSGGLVCLVGS 240
181 lavpidtpkcnnllyskdtfegyqptikndmlcagfeegkdkacgdsqgplvclvgs 240
241 WLQAGVTSNMGEGCARONRPGVYTRVTAHNMWIRHRIIPKLOFORARLGOK 290
241 wlqagvtsnmgegcaronrpgvytrvtahnmwirhriipklofparlgok 290

RESULT 4

AAE03945 standard; Protein; 290 AA.

AAE03945:

29-MAY-2001 (first entry)

| | CC | assays for receptor activity, arthritis and inflammation, leukemias and C.N.S disorders. |
|--------|---|--|
| CC | C.C.N.S disorders. | |
| CC | Note: The sequence data for this patent did not form part of the printed specification. | |
| CC | | |
| xx | Sequence | 328 AA; |
| SO | | |
| | Query Match | 97.3%; Score 1528; DB 22; Length 328; |
| | Best Local Similarity | 97.9%; Pred. No. 1.2e-130; |
| | Matches 281; Conservative | 3; Mismatches 3; Indels 0; Gaps 0; |
| OY | 1 | MRRPAVLLLLLLLCGSGRAKATCGRRPMINRWVGSDODTEGGMPQVSIQKNGSHFC 60 |
| DB | 30 | mrrpaavlllllclfgsqrakatacagprmlnmhvsgqddlegawpvsifqngshfc 89 |
| OY | 61 | GGSILAEOMWVLTAAICFNTSETSLYOVLLGAROLVOPGRNMYARVQVESNPPLYOGTA 120 |
| DB | 90 | ggsliaeqwvllaacfnfntsetsllyqvllgarqlvpgrphmyarvtqvvesnplyqgta 149 |
| OY | 121 | SSADVALVELAEAPVFPTNYILPVCCLPDPSVILETGMNCWVTGWSGSPSEEDLPEPRILQK 180 |
| DB | 150 | ssadvalvelaeapvfptnyilpvcclpdpsviletgmmncwvtgwspsedl1peprilqk 209 |
| OY | 181 | LAVPIIDPKCNLLYSKOTEGYORKTKNDMLCAGFEBGKKDACKGDGSGEPLCVLGQS 240 |
| DB | 210 | lavpiidpcrnllyskotefgyorktklndmlcagfeegkkdackgdgaaplvc1vgqs 269 |
| OY | 241 | WLQACVISMGSCGACRONPGVYIRTAHNHNRILIPKLOPORALIG 287 |
| DB | 270 | wlqavgvismgscargcarqpvyilrtahnhwlrhlrlpk1qvqpsavg 316 |
| RESULT | 7 | |
| ID | AAM77304 | |
| XX | AAM77304 standard; Protein: | 297 AA. |
| XX | AAM77304; | |
| DT | 07-JAN-1999 | (first entry) |
| XX | | |
| DE | Amino acid sequence of SP0033LA, a homologue of HELA2. | |
| KX | Serine protease; regulation; cell activity; viability; HELA2; ATC2; | |
| KW | BCOM3; testisin; fertility; suppressor; testicular germ cell cancer; | |
| KM | semnoma; testis-specific expression; antitumour; sperm development; | |
| KW | interfertility; human; chromosome 10p13.3. | |
| XX | Homo sapiens. | |
| OS | | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | Disulfide-bond 1 | /note= "likely to be involved in disulphide bonding" |
| FT | Disulfide-bond 36 | /note= "likely to be involved in disulphide bonding" |
| FT | Misc-difference 51 | /note= "likely to be a catalytic residue" |
| FT | Disulfide-bond 52 | /note= "likely to be involved in disulphide bonding" |
| FT | Misc-difference 100 | /note= "likely to be involved in disulphide bonding" |
| FT | Disulfide-bond 134 | /note= "likely to be a catalytic residue" |
| FT | Disulfide-bond 167 | /note= "likely to be involved in disulphide bonding" |
| FT | Disulfide-bond 190 | /note= "likely to be involved in disulphide bonding" |
| FT | Disulfide-bond 201 | /note= "likely to be involved in disulphide bonding" |
| FT | Misc-difference 205 | /note= "likely to be involved in disulphide bonding" |
| FT | Disulfide-bond 211 | /note= "likely to be a catalytic residue" |
| FT | Disulfide-bond 211 | /note= "likely to be involved in disulphide bonding" |

RESULT 9
 AAM8453
 ID AAM8453 standard; Protein; 304 AA.
 XX
 AC AAM8453;
 XX
 DT 10-MAY-1999 (first entry)
 XX
 DE Serine protease HE2NM40.
 XX
 KM HE2NM40; serine protease; human; cancer; inflammation; asthma;
 KM wasting disease; atherosclerosis; stroke; diabetes; arthritis;
 KM neurodegenerative disease; Alzheimer's disease; diagnosis;
 KM therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 PF EP890646-A2.
 XX
 PR 13-JAN-1999.
 XX
 PF 01-JUN-1998; 98EP-0304312.
 XX
 PR 20-FEB-1998; 98GB-0003650.
 PR 10-JUN-1997; 97GB-0012088.
 PR 17-OCT-1997; 97EP-0308295.
 XX
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Burgess NA, Clinkenbeard HE, Southan CD;
 XX
 DR WPI: 1999-072882/07.
 DR N-PSDB: AAX06942.
 XX
 PT New serine protease HE2NM40 polypeptides and polynucleotides -
 PT useful as diagnostic reagents and for prevention and treatment of
 PT cancer, Alzheimer's disease and asthma
 XX
 PS Claim 1; Page 16-17; 21pp; English.
 XX
 CC This polypeptide, termed HE2NM40, comprises a serine protease that
 CC that shows homology to human trypsin. Its amino acid sequence was
 CC deduced from an isolated cDNA clone (see AAX06942). HE2NM40
 CC polypeptides and polynucleotides are useful for diagnosing diseases
 CC mutations in the HE2NM40 gene, and/or analysing the presence or
 CC amount of expressed polypeptide (claimed). HE2NM40 polypeptides,
 CC polynucleotides (gene therapy), antisense sequences, agonists and
 CC antagonists, and soluble polypeptides that bind the HE2NM40 ligand
 CC to prevent expression, can be used to modulate HE2NM40 expression
 CC or activity. Diseases diagnosed, prevented or treated include
 CC cancer, inflammation, asthma, wasting diseases, atherosclerosis,
 CC stroke, diabetes, arthritis, and neurodegenerative diseases and
 CC disorders including Alzheimer's disease. HE2NM40 antibodies are
 CC useful for inducing an immune response to immunise and prevent
 CC disease, and for isolating HE2NM40 clones or purifying the
 CC polypeptide by affinity chromatography. HE2NM40 polypeptides can
 CC be administered directly or as a vaccine to inoculate against
 CC disease.
 XX
 SQ Sequence 304 AA;

Query Match 76.28; Score 1197; DB 20; Length 304;
 Best Local Similarity 82.14; Pred. No. 1.3e-100;
 Matches 234; Conservative 7; Mismatches 36; Indels 8; Gaps 3;

QY 1 MRRRAAVPLLLLLCGSGORAKATACGRPTLNMVGGCOTQGESEWQVSIQNGSHFC 60
 Db 1 MRRRAAVPLLLLLCGSGORAKATACGRPTLNMVGGCOTQGESEWQVSIQNGSHFC 60
 QY 61 GGSIIADQWLTAAHCRNMTSETSLYCVLLGARQLVPGHAMARVQVESNPLVCGTA 120
 Db 61 GGSIIADQWLTAAHCRNMTSETSLYCVLLGARQLVPGHAMARVQVESNPLVCGTA 120

Db 61 ggsIIaegwltaaHcfrntsetsldqvl]garg]vqpgphamyarvqvesnplygta 120
 QY 121 SSADVALVELAPVPEFTNYILPVLDPSPVYFETGMNCWYGMGSPSEEDLLPEPRILQ 180
 Db 121 ssayvalvelapvpeftnyilpvlclpdqgrlftetgmncwytgmspsedllpeprilq 180
 QY 181 LAVPIIDPKCNLYLSKDTFEGYQPKTIKNDMLCAGFECKKADKSGGPLVCLVGS 240
 Db 181 lavpidtpkcnyllskdtfeqyqpktkndmlcagl--srrarmpasgagmcmqld 238
 QY 241 WLQAGV-LSWGEG-----CARQNRPGYIRYTAHNMNTHRIPTL 279
 Db 239 fqrpgpggvglsprywmqltpepsvllrvngqgmwlhriplxl 283

RESULT 10
 AAY06482
 ID AAY06482 standard; Protein; 317 AA.
 XX
 AC AAY06482;
 XX
 DT 27-SEP-1999 (first entry)
 XX
 DE Human tumour-associated protein PRO343.
 XX
 KM PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT Peptide 1..32
 FT Protein /note="signal peptide"
 FT Modified-site 33..317
 FT /note="mature protein"
 FT /note="N-glycosylated"
 XX
 PN WC935170-A2.
 PD 15-JUL-1999.
 PF 05-JAN-1999; 99MO-US00106.
 XX
 PR 20-NOV-1998; 98US-0109304.
 PR 05-JAN-1998; 98US-0070440.
 PR 29-APR-1998; 98US-0083500.
 PR 22-MAY-1998; 98US-0086414.
 PR 10-JUN-1998; 98US-0088742.
 PR 10-NOV-1998; 98US-0107783.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;
 PI Roy MA, Wood WI;
 XX
 DR WPI: 1999-430385/36.
 DR N-PSDB: AAX87259.
 XX
 PT Antibody against proteins expressed in neoplastic cells, useful for
 PT tumor diagnosis and treatment
 XX
 PS Example 1; Fig 12; 162pp; English.
 XX
 CC This sequence represents human PRO343 (UNQ302), a protein encoded
 CC by the novel cDNA clone DNA43318 (see AAX87259). Amplification of
 CC DNA43318 was observed in primary lung tumours and in primary colon
 CC tumours, suggesting a significant role in tumour formation and
 CC growth. Antagonists (e.g. antibodies) directed to PRO343 may have
 CC use in cancer therapy. The invention identifies 14 genes (see
 CC AAX87254-67) that are amplified in the genome of tumour cells. Such
 CC amplification is expected to be associated with overexpression of
 CC the gene product and to contribute to tumorigenesis. The encoded
 CC proteins (see AAY06477-90) may be useful targets for the diagnosis

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:56:23 ; Search time 30.25 Seconds
(without alignments)
1658.464 Million cell updates/sec

Title: US-10-041-006-7

Perfect score: 1571
Sequence: 1 MRRPAAVLLLLLCRCGQRA.....WTHRIPLKQFQPARLGQK 290

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_19:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.rvivirus:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 721.5 | 45.9 | 297 | 11 | 088781 rat |
| 2 | 682.5 | 43.4 | 389 | 13 | 09PVX7 xenopus lae |
| 3 | 677.5 | 43.1 | 317 | 13 | 09DGR3 xenopus lae |
| 4 | 633.5 | 40.3 | 310 | 11 | 091XC4 mus musculus |
| 5 | 623.5 | 39.7 | 310 | 11 | 09QY29 mus musculus |
| 6 | 612.5 | 39.0 | 321 | 4 | 096R28 mus musculus |
| 7 | 610.5 | 38.9 | 339 | 11 | 099144 mus musculus |
| 8 | 583.5 | 37.1 | 273 | 11 | 0921N4 mus musculus |
| 9 | 573.5 | 36.5 | 275 | 4 | 096R26 mus musculus |
| 10 | 555 | 35.3 | 799 | 11 | 09DB10 mus musculus |
| 11 | 548 | 34.9 | 643 | 6 | 097506 sus scrofa |
| 12 | 545.5 | 34.7 | 273 | 6 | 09XSM1 ovis aries |
| 13 | 542.5 | 34.5 | 329 | 13 | 042272 xenopus lae |
| 14 | 525 | 33.4 | 237 | 6 | 046507 papio hamad |
| 15 | 513.5 | 32.7 | 237 | 6 | 029464 bos taurus |
| 16 | 505 | 32.1 | 806 | 6 | 018783 macropus eu |

| | | | | | |
|----|-------|------|------|----|---------------------|
| 17 | 503 | 32.0 | 454 | 6 | 046506 papio hamad |
| 18 | 503 | 32.0 | 810 | 4 | 015146 mus musculus |
| 19 | 500.5 | 31.9 | 812 | 11 | 09R0W3 mus musculus |
| 20 | 498.5 | 31.7 | 322 | 11 | 0920S2 mus musculus |
| 21 | 494.5 | 31.5 | 490 | 11 | 0920K3 mus musculus |
| 22 | 486.5 | 31.0 | 581 | 4 | 09BIE2 mus musculus |
| 23 | 486 | 30.9 | 282 | 11 | 09DA13 mus musculus |
| 24 | 485.5 | 30.9 | 624 | 6 | 095ME7 mus musculus |
| 25 | 484 | 30.8 | 267 | 5 | 09BK47 mus musculus |
| 26 | 483 | 30.7 | 624 | 11 | 09DA13 mus musculus |
| 27 | 481.5 | 30.6 | 812 | 11 | 091W5 mus musculus |
| 28 | 478.5 | 30.5 | 537 | 4 | 09BIE1 mus musculus |
| 29 | 477 | 30.4 | 624 | 11 | 091Y47 mus musculus |
| 30 | 471 | 30.0 | 492 | 4 | 096T73 mus musculus |
| 31 | 469 | 29.9 | 1524 | 13 | 091674 mus musculus |
| 32 | 467.5 | 29.8 | 405 | 4 | 096E86 mus musculus |
| 33 | 462.5 | 29.4 | 855 | 11 | 09JIT7 mus musculus |
| 34 | 462 | 29.4 | 329 | 6 | 09GL10 mus musculus |
| 35 | 455 | 29.0 | 467 | 5 | 0967X8 mus musculus |
| 36 | 454.5 | 28.9 | 279 | 11 | 099MS4 mus musculus |
| 37 | 453 | 28.8 | 326 | 11 | 09D9M0 mus musculus |
| 38 | 452.5 | 28.8 | 263 | 11 | 09CR35 mus musculus |
| 39 | 450.5 | 28.7 | 263 | 11 | 09DC86 mus musculus |
| 40 | 449.5 | 28.6 | 845 | 13 | 09DGR1 mus musculus |
| 41 | 448 | 28.5 | 279 | 11 | 09QZ74 mus musculus |
| 42 | 448 | 28.5 | 415 | 6 | 029015 mus sp. pre |
| 43 | 446.5 | 28.4 | 260 | 13 | 09W7O3 mus musculus |
| 44 | 446.5 | 28.4 | 263 | 11 | 09DX8 mus musculus |
| 45 | 445.5 | 28.4 | 767 | 13 | 09DGR2 mus musculus |

ALIGNMENTS

| RESULT | ID | PRELIMINARY | PRT | 297 AA |
|--------|---|-------------|-----|--------|
| 1 | 088781 | | | |
| AC | 088781 | PRELIMINARY | PRT | 297 AA |
| DT | 01-NOV-1998 (TREMBLrel. 08, Created) | | | |
| DT | 01-NOV-1998 (TREMBLrel. 08, Last sequence update) | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Last annotation update) | | | |
| DE | SERINE PROTEASE PRECURSOR (FRAGMENT) | | | |
| GN | BSP2. | | | |
| OS | Rattus rattus (Black rat). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | |
| OX | NCBI_TaxID=10117; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=FISHER; TISSUE=BRAIN; | | | |
| RC | MDLINE=98389725; PubMed=9722524; | | | |
| RA | Davies B.J., Pickard B.S., Steel M., Morris R.G., Lath R.; | | | |
| RT | "Serine Proteases in Rodent Hippocampus"; | | | |
| RU | J. Biol. Chem. 273:23004-23011(1998) | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | | |
| CC | TRYPSIN FAMILY. | | | |
| DR | EMBL; AJ005642; CAA0644.1; - | | | |
| DR | HSSP; P00763; IDPO. | | | |
| DR | MEROPS; S01.252; - | | | |
| DR | InterPro; IPR001314; Chymotrypsin. | | | |
| DR | InterPro; IPR001254; Trypsin. | | | |
| DR | Pfam; PF00899; Trypsin_1. | | | |
| DR | PRINTS; PR00722; CHYMOTRYPSIN. | | | |
| DR | SMART; SM00020; TRYP_SP; 1. | | | |
| DR | PROSITE; PS50240; TRYPSIN_DOM; 1. | | | |
| DR | PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1. | | | |
| DR | PROSITE; PS00135; TRYPSIN_SER; 1. | | | |
| KW | Hydrolase; Protease; Serine protease; Signal. | | | |
| FT | NON_TER | | | |
| FT | SIGNAL | | | |
| FT | CHAIN | | | |
| FT | SEQUENCE | | | |


```

      | | : | : : |
Db 268 YTFVPAYETWISERSV--ISFKP 288

```

| | | |
|--------|--------|--------------|
| RESULT | 4 | |
| Q91XC4 | | |
| ID | Q91XC4 | PRELIMINARY; |
| ID | Q91XC4 | PRT; 310 AA |

DT 01-DEC-2001 (TREMBLrel . 19, created)
 DT 01-DEC-2001 (TREMBLrel . 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel . 19, last annotation update)
 DE SIMILAR TO DISTAL INTESTINAL SERINE PROTEASE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
 NCBI_TaxID=10090;

| | | | | |
|-----------------------|-----------------|-------------------|----------|------------|
| Query Match | 40.38% | Score 633.5 | DB 11 | Length 310 |
| Best Local Similarity | 46.48% | Pred. No. 1.1e-54 | | |
| Matches 129 | Conservative 37 | Mismatches 103 | Indels 9 | Gaps 6 |

[illegible]

```

RESULT      5
090xyz29
ID 090xyz29 PRELIMINARY; PRT; 310 AA.
AC 090xyz29;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE DISTAL INTESTINAL SERINE PROTEASE.
GN DISP OR DISP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20246293; Pubmed=10786627;
RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,
RT Walters J.R.;
RA "Characterization of a novel murine intestinal serine protease,
RT DISP. ";
RL Biochim. Biophys. Acta 1490:131-136(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.
DR EMBL; AJ243866; CAB56465.1; -
DR HSP; P00763; 1DPO.
HSP; P00763; 1DPO.

DR MGd; MG1.1353645; Dusp.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR01254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYP-SIN.
DR SMART; SM00020; TRY-SPEC; 1.
DR PROSITE; PS02404; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CMC64;

| | | | | | | | |
|-----------------------|-------|--------------|-------------|------------|-----|--------|-----|
| Query Match | 39.7% | Score | 623.5 | DB | 11 | Length | 310 |
| Best Local Similarity | 46.0% | Pred. | No. 1.1e-53 | | | | |
| Matches | 128 | Conservative | 36 | Mismatches | 105 | Indels | 9 |
| | | | | | | Gaps | 6 |

```

QY 3 RRAAAVPLLLLLCFCGSOAK -AAVACGKRRKRRNRKGGGJVLGGEMFYQS 1VANSJH
Db 4 RARCFLLLLLIITLRARADILPEVCGSHRAGKIVGGQDALEGGMPQVMSLMTEDS -RI 62
QY 60 CGGSLIAECQWVLTAAHCFRRNTSETSLYOVLGAGALVOPGPHAAVARYAROVESNPLY -QG 118
Db 63 CGGSLIHEWVLTAAHCFRRSRLNSFYHVKKGGLTLLSEPHSLVAVRNTFVHPITLMA 122
QY 119 TASSADVALVELAEAVPFTNYILPVCGLDPDPVIEFTCGMNCVATGMSGPSEEDILPEPRIL 178
Db 123 DASSDDIALVQDPLRRLRSQF -TPVCEPAAQTPLPPLGVCVAVTGMGATQERDM ---ASVL 178
QY 179 QKLAAPILIDPRPCNLXSKDREFVQOPTKINDMCAAGEEGKKDACKDGGSGGPLYWG 238
Db 179 QELAAPFLDESECEKMYHTQGSSLSGERLITQSDMLCAGIYVGHGHI DSCGDSGGPVCYSIN 238
QY 239 QSMQLAGVITSMGEGCARONRPGVYTRVTAHHNHITRI 276
Db 239 SSMTOVGITSMGIGCARPYRPGVYTRVTAHYDMIDORIL 276

```

| | |
|--------|--|
| RESULT | 6 |
| Q96R28 | |
| ID | Q96R28; PRELIMINARY; PRT; 321 AA. |
| AC | Q96R28; |
| DT | 01-DEC-2001 (TREMblrel_19, Created) |
| DT | 01-DEC-2001 (TREMblrel_19, last sequence update) |
| DT | 01-DEC-2001 (TREMblrel_19, last annotation update) |
| DE | HS TRANSMEMBRANE TRYPASE, GENE NAME TMT_A175522.1. |

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:21096910; PubMed-11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tuftarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
DR EMBL; AF006466; AAK61269.1; -.
KW Transmembrane;
SQ SEQUENCE 321 AA; 33829 MW; FFF5089ED0C4FE73D CRC64;

SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

SQ SEQUENCE 321 AA; 33829 MW; FFF5089EDC4FC73D CRC64;

| | | | | | | | |
|-----------------------|--------|--------------|----------|------------|-----|--------|------|
| Query Match | 39.0%; | Score | 612.5; | DB | 4; | Length | 321; |
| Best Local Similarity | 47.3%; | Pred. No. | 1.5e-52; | | | | |
| Matches | 129; | Conservative | 36; | Mismatches | 91; | Indels | 17; |
| | | | | | | Gaps | 7 |

QY 9 LLLLCGSGORAKATACGRRPMLN---RMVGGODTQEGEMPMQVSIORNGSHFGCGSLI 65
 DB 9 LLLLAAGVGLRTLPQCGRRQVSDAGRGIVGHAARACAMPMQASIRLRKRVYCGGSLI 68
 QY 66 AEGMVLTAACFRMTSETSLYVLLGARQLVQRPAMARARVQV---ESNPLOYGTASSA 123
 DB 69 SPQWVLTAACFGSLNSDSDYVHLGELITL-SPH--FSTVROILHSSPSGO-PGTSG 124
 QY 124 DVALVELAPVPTNYILPVCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILQKAV 183
 DB 125 DIALVELSVPTLSSRLPVCLEASDDFCGIRCVWVGTRBGEPLRPPLYSLEKRV 184
 QY 184 PIIDTPKCNLLYSKDTFEGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSQW 243
 DB 185 SYVDTEFCRRDYP-----GPGGSIILQPMICA---RGPQDAGDSDGGLVQVGNAMW 236
 QY 244 AGVISMGEGCARQNRPGVYIRVTAHHNMTIRRI 276
 DB 237 AGVISMGEGCARQNRPGVYIRVTAHHNMTIRRI 269
 7
 ID 099144 PRELIMINARY: PRT: 339 AA.
 AC 099144: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE SIMILAR TO PROTEASE, SERINE, 8 (PROSTASIN).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV, AND SWISS; TISSUE=LUNG;
 RA Verghese G.M., Caughey G.H.;
 RT "Molecular cloning and characterization of mouse prostasin, a type I
 RT membrane-associated serine protease of the gamma-tryptase/prostasin
 RL gene family.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL: BC003851; AA03851.1; -;
 DR EMBL: AF378086; AAL06320.1; -;
 DR EMBL: AF378085; AAL06319.1; -;
 DR EMBL: P00734; I0VS.
 DR MEROPS: S01.158; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP-SPC; 1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 339 AA; 36216 MW; BC2DE88BC057AF10 CRC64;

Query Match 38.9%; Score 610.5; DB 11; Length 339;
 Best Local Similarity 42.3%; Pred. No. 2,5e-52;
 Matches 119; Conservative 48; Mismatches 103; Indels 11; Gaps 4;

QY 6 AVPLLLILCF---GSGRAKATACG---RPMRLMVGODTQEGEMPMQVSIORNGSHF 59
 DB 14 AVTLLGLGSGIRADCTEASCAGVIOF---RITGGGSAKPGQPMQVSIYIDGNHV 69

QY 60 CGGSLIAEGMVLTAACFRMTSETSLYVLLGARQLVQRPAMARARVQV---ESNPLOYGT 119
 DB 70 CGGSLVSNKMWVYSAACFRPREHREAREVYKLGARHOLDYSNDTVVHTVAQIITHSYREE 129
 QY 120 ASSADVAVELEAPVPTNYILPVCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILQ 179
 DB 130 GSGGIALIRLSSPTFSRTYRIPCLPANAASFNGLCCTYTGWGVAPSVLQTPRLQ 189
 QY 180 KIAVDPIIDTPKCNLLYSKDTFEGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSQW 239
 DB 190 QLEVPLISRETSCSLYININA-VPEEPHTIQQDMICAGYKVGKDGKACGGDGLSCPMEG 248
 QY 240 SWLAGVISMGEGCARQNRPGVYIRVTAHHNMTIRRI 280
 DB 249 IWTLAGVISMGEGCARQNRPGVYIRVTAHHNMTIRRI 289
 RESULT 8
 ID 0921N4 PRELIMINARY: PRT: 273 AA.
 AC 0921N4: 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE MAST CELL PROTEASE 7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC011528; AA011528.1; -;
 KW Protease.
 SQ SEQUENCE 273 AA; 30332 MW; 51417476514035BE CRC64;

Query Match 37.1%; Score 583.5; DB 11; Length 273;
 Best Local Similarity 45.1%; Pred. No. 8.9e-50;
 Matches 123; Conservative 42; Mismatches 99; Indels 9; Gaps 5;

QY 10 LLLLCGSGORAKATACGRRPMLNRMVGGODTQEGEMPMQVSIORNGSHFGCGSLI 66
 DB 4 LLLTLPLLSVLAAAGPAMTREGIVGGEAHGNKMPQVSIKANDYTMHFCGSLIH 63
 QY 67 EGMVLTAACFRMTSETSLYVLLGARQLVQRPAMARARVQV---ESNPLOYGTASSADY 125
 DB 64 PGMVLTAACFRMTSETSLYVLLGARQLVQRPAMARARVQV---ESNPLOYGTASSADY 120
 QY 126 ALVELAPVPTNYILPVCLPDPSPVIFETGMNCWVTGSGSPSEEDLLPERILQKAV 185
 DB 121 ALKLTMPVVISDYVIRVPLRPASSETPSTGLCWVGMGINDGVNLPPLPKKEYOVP 180
 QY 186 IDTPKCNLLYSKDTFEGYQPKTKNDMLCAGFEKGKDKGSGGPLVCLVGSQW 245
 DB 181 IENHLCDLKYKGLITGDVNIIVRDMCLAG-NEG-HDSCGDSGGGPLVCKVEDMTLQAG 238
 QY 246 VISMGEGCARQNRPGVYIRVTAHHNMTIRRI 278
 DB 239 VISMGEGCARQNRPGVYIRVTAHHNMTIRRI 271

RESULT 9
 ID 096RZ6 PRELIMINARY: PRT: 275 AA.
 AC 096RZ6: 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
 DE TRYPTASE 1.
 GN Homo sapiens (Human).

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: August 13, 2002, 08:57:03 ; Search time 13.47 Seconds
(without alignments)
833.606 Million cell updates/sec

Title: US-10-041-006-7

Perfect score: 1571
Sequence: 1 MRPAVPLLLLCFSGSRA.....WIRRIPLKQFQPARLGQK 290

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 1571 | 100.0 | 290 | 1 MPN_HUMAN | O9gqz3 homo sapien |
| 2 | 739 | 47.0 | 317 | 2 BSS4_HUMAN | O9gzn4 homo sapien |
| 3 | 696 | 44.3 | 306 | 1 BSS4_MOUSE | O9er10 mus musculu |
| 4 | 618.5 | 39.4 | 343 | 1 PSS8_HUMAN | O16651 homo sapien |
| 5 | 614.5 | 39.1 | 342 | 1 PSS8_RAT | O9es87 rattus norv |
| 6 | 613.5 | 39.1 | 321 | 1 TRYG_HUMAN | O9nrr2 homo sapien |
| 7 | 610.5 | 38.9 | 311 | 1 TRYG_MOUSE | O9qu17 mus musculu |
| 8 | 606.5 | 38.6 | 342 | 1 PSS8_MOUSE | O9esd1 mus musculu |
| 9 | 594 | 37.8 | 270 | 1 TRYT_MERUN | O9esd1 mus musculu |
| 10 | 590.5 | 37.6 | 273 | 1 MCT7_RAT | P20231 homo sapien |
| 11 | 588.5 | 37.5 | 275 | 1 TRB1_HUMAN | O15661 mus musculu |
| 12 | 587.5 | 37.4 | 275 | 1 TRB1_MOUSE | O9nrd1 sus scrofa |
| 13 | 584.5 | 37.2 | 273 | 1 MCT7_MOUSE | O02844 mus musculu |
| 14 | 578.5 | 36.8 | 273 | 1 TRYT_MOUSE | O9yem0 homo sapien |
| 15 | 574 | 36.5 | 314 | 1 TRYT_HUMAN | P15157 homo sapien |
| 16 | 571.5 | 36.4 | 275 | 1 TRYT_HUMAN | O50343 rattus norv |
| 17 | 565.5 | 36.0 | 274 | 1 MCT6_RAT | P21845 mus musculu |
| 18 | 564 | 35.9 | 276 | 1 MCT6_MOUSE | P15944 mus musculu |
| 19 | 560.5 | 35.7 | 275 | 1 TRYT_MOUSE | O9jnj7 sus musculu |
| 20 | 559.5 | 35.6 | 324 | 1 TRYT_MOUSE | O9xsm2 ovis aries |
| 21 | 556.5 | 35.4 | 273 | 1 TRYT_SHEEP | P03952 homo sapien |
| 22 | 533 | 33.9 | 638 | 1 KAL_HUMAN | P12545 macaca mula |
| 23 | 526 | 33.5 | 638 | 1 PLMN_MACMU | O35453 mus musculu |
| 24 | 523.5 | 33.3 | 638 | 1 HEP5_MOUSE | P26262 mus musculu |
| 25 | 518 | 33.0 | 638 | 1 KAL_MOUSE | P05981 mus musculu |
| 26 | 517.5 | 32.9 | 417 | 1 HEP5_HUMAN | P14272 homo sapien |
| 27 | 517 | 32.9 | 638 | 1 KAL_RAT | P80010 equus cabal |
| 28 | 509 | 32.4 | 338 | 1 PLMN_HORSE | O05511 rattus norv |
| 29 | 505.5 | 32.2 | 416 | 1 HEP5_RAT | O9et04 mus musculu |
| 30 | 504.5 | 32.1 | 810 | 1 TMS5_MOUSE | P00747 homo sapien |
| 31 | 502 | 32.0 | 812 | 1 PLMN_HUMAN | P06868 bos taurus |
| 32 | 496.5 | 31.6 | 333 | 1 PLMN_BOVIN | P80009 canis faml1 |
| 33 | 496 | 31.6 | 333 | 1 PLMN_CANFA | |

| | | | | | |
|----|-------|------|------|--------------|--------------------|
| 34 | 493 | 31.4 | 269 | 1 TRYM_CANFA | P19236 canis faml1 |
| 35 | 491.5 | 31.3 | 1034 | 1 ENTK_PIG | P98074 sus scrofa |
| 36 | 488.5 | 31.1 | 450 | 1 TMS2_MOUSE | O91jg8 mus musculu |
| 37 | 487.5 | 31.0 | 625 | 1 FALL_HUMAN | P03951 homo sapien |
| 38 | 485.5 | 30.9 | 457 | 1 TMS2_HUMAN | O9n3s3 homo sapien |
| 39 | 481 | 30.6 | 422 | 1 DES1_HUMAN | O9u1s2 homo sapien |
| 40 | 479.5 | 30.5 | 812 | 1 PLMN_MOUSE | P20918 mus musculu |
| 41 | 479 | 30.5 | 790 | 1 PLMN_PIG | P06867 sus scrofa |
| 42 | 478 | 30.4 | 343 | 1 PLMN_SHEEP | P1417 macaca mula |
| 43 | 477.5 | 30.4 | 1420 | 1 AEON_MACMU | P1417 macaca mula |
| 44 | 476.5 | 30.3 | 855 | 1 ST14_HUMAN | O9y5y6 homo sapien |
| 45 | 471 | 30.0 | 492 | 1 TMS2_HUMAN | O13593 homo sapien |

ALIGNMENTS

| RESULT | ID | MPN_HUMAN | STANDARD | PRT | 290 AA. |
|--------|--|-----------------------------------|-----------|-------------------------|--------------------------------------|
| AC | O9BOR3 | 16-OCT-2001 (Rel. 40, Created) | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last sequence update) | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last annotation update) | | | |
| DE | Marapsin precursor (EC 3.4.21.-) | | | | |
| GN | MPN | | | | |
| OS | Homo sapiens (Human) | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | |
| OX | NCBI_Taxid=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.; | | | | |
| RT | Cloning, sequencing and expression of marapsin, a human serine | | | | |
| RL | protease. | | | | |
| CC | Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | | | |
| CC | TRYPsin FAMILY. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation- | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@sib-sib.ch). | | | | |
| CC | ----- | | | | |
| CC | EMBL: AJ306593; CAC35467.1; - | | | | |
| CC | MEROPS: S01.074; - | | | | |
| CC | InterPro: IPR001314; Chymotrypsin. | | | | |
| CC | InterPro: IPR001234; Trypsin. | | | | |
| CC | Pfam: PF00089; Trypsin; 1. | | | | |
| CC | PRINTS: PR00722; CHYMOTRYPSIN. | | | | |
| CC | SMART: SM00020; TRYP-SPC; 1. | | | | |
| CC | PROSITE: PS0240; TRYPsin; 1. | | | | |
| CC | PROSITE: PS00134; TRYPsin; HIS; 1. | | | | |
| CC | DR PROSITE: PS00135; TRYPsin; SER; 1. | | | | |
| CC | KW Hydrolyse; Serine protease; Lysogen; Signal; Glycoprotein. | | | | |
| CC | FT SIGNAL | 1 | 22 | | POTENTIAL. |
| CC | FT PROPEP | 23 | 34 | | POTENTIAL. |
| CC | FT CHAIN | 35 | 290 | | MARAPsin. |
| CC | FT DOMAIN | 35 | 277 | | SERINE PROTEASE. |
| CC | FT ACT_SITE | 75 | 75 | | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| CC | FT ACT_SITE | 124 | 124 | | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| CC | FT ACT_SITE | 229 | 229 | | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| CC | FT DISULFID | 60 | 76 | | BY SIMILARITY. |
| CC | FT DISULFID | 158 | 235 | | BY SIMILARITY. |
| CC | FT DISULFID | 191 | 214 | | BY SIMILARITY. |
| CC | FT DISULFID | 225 | 253 | | BY SIMILARITY. |
| CC | FT CAROHD | 55 | 55 | | N-LINKED (GLCNAC...) (POTENTIAL). |
| CC | FT CAROHD | 79 | 79 | | N-LINKED (GLCNAC...) (POTENTIAL). |
| CC | SEQUENCE | 290 AA; | 31940 MW; | 67BDC93BC70BFF7B CRC64; | |

Query Match 100.0%; Score 1571; DB 1; Length 290;
 Best Local Similarity 100.0%; Pred. No. 1,3e-133;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRPAVPLLLLCFSGSORAKAATACGRPMILNRMGVGGDTGEGMPVOSTORNSHFC 60
 DB 1 MRRPAVPLLLLCFSGSORAKAATACGRPMILNRMGVGGDTGEGMPVOSTORNSHFC 60
 QY 61 GGSIIAEOWVLTAAHCFRNTSETSLYOVLAGAQLVQPGPHAMTARVQESNPLYQSTA 120
 DB 61 GGSIIAEOWVLTAAHCFRNTSETSLYOVLAGAQLVQPGPHAMTARVQESNPLYQSTA 120
 QY 121 SSADVALLVLEAPVFTNYILPVCLPDPVSIFETGMNCWVTGSGSPSEDLPEPRILQK 180
 DB 121 SSADVALLVLEAPVFTNYILPVCLPDPVSIFETGMNCWVTGSGSPSEDLPEPRILQK 180
 QY 181 LAVPIIDPKCNILYSKDFEQPTKINDMLCAGFEFGKDKGSGGLVCLVQSGS 240
 DB 181 LAVPIIDPKCNILYSKDFEQPTKINDMLCAGFEFGKDKGSGGLVCLVQSGS 240
 QY 241 WLAGVISMWEGCARONRPGVYIRVTAHNMHRIIPKLOFOPARLGQK 290
 DB 241 WLAGVISMWEGCARONRPGVYIRVTAHNMHRIIPKLOFOPARLGQK 290

RESULT 2
 BSS4_HUMAN STANDARD; PRT; 317 AA.
 ID BSS4_HUMAN 04342;
 AC 09GZM4; 04342;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4) (SP001LA).
 GN PRSS26 OR PRSS22 OR BSSP4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Mitsui S., Okui A., Komiami K., Yamaguchi N.;
 RT "Cloning and characterization of a human brain-specific serine protease, hbssp-4.";
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Pancreas;
 RA Young G.W., Stevens R.L.;
 RT Identification of a new member of the chromosome 16 family of serine proteases.";
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 47-317 FROM N.A.
 RA Riche D.O., Bruce D., Mundt N., Doggett N., Munk C., Saunders E., Robinson D., Jones M., Buckingham J., Chastain L., Thompson S., Goodwin L., Bryant J., Resmer J., Melnick L., Longstre J., White S., Veng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
 RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; AB010779; BAB20263.1; -

DR EMBL; AF321182; AAG35070.1; -;
 DR EMBL; AC003965; AAB93671.1; -;
 DR MEROPS; S01.252; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP-Spc; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 317
 FT ACT_SITE 90 90
 FT ACT_SITE 141 141
 FT ACT_SITE 242 242
 FT DISULFID 75 91
 FT DISULFID 175 248
 FT DISULFID 208 227
 FT DISULFID 238 266
 FT CARBOHYD 70 70
 FT CONFLICT 47 47
 FT SEQUENCE 317 AA; 33731 MW; E2A123BC66E79935 CRC64;

Query Match 47.0%; Score 739; DB 1; Length 317;
 Best Local Similarity 48.8%; Pred. No. 6.3e-60;
 Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;

QY 10 LLLLCFSGSORAKAATACGRPMILNRMGVGGDTGEGMPVOSTORNSHFCGSLI 65
 DB 21 LLLLCFSGSORAKAATACGRPMILNRMGVGGDTGEGMPVOSTORNSHFCGSLI 65
 QY 66 AEWVLTAAHCFRNTSETSLYOVLAGAQLVQPGPHAMTARVQESNPLYQSTA 123
 DB 81 TSWVLTAAHCFRNTSETSLYOVLAGAQLVQPGPHAMTARVQESNPLYQSTA 123
 QY 124 DVALLVLEAPVFTNYILPVCLPDPVSIFETGMNCWVTGSGSPSEDLPEPRILQK 183
 DB 141 DVALLVLEAPVFTNYILPVCLPDPVSIFETGMNCWVTGSGSPSEDLPEPRILQK 183
 QY 184 PLIDPKCNILYSKDFEQPTKINDMLCAGFEFGKDKGSGGLVCLVQSGSLQ 243
 DB 201 PLIDPKCNILYSKDFEQPTKINDMLCAGFEFGKDKGSGGLVCLVQSGSLQ 243
 QY 244 AGVISMWEGCARONRPGVYIRVTAHNMHRIIPKLOFOPARLGQK 288
 DB 257 AGVISMWEGCARONRPGVYIRVTAHNMHRIIPKLOFOPARLGQK 288

RESULT 3
 BSS4_MOUSE STANDARD; PRT; 306 AA.
 ID BSS4_MOUSE 04342;
 AC 09GZM4; 04342;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4).
 GN PRSS26 OR BSSP4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Mitsui S., Okui A., Komiami K., Yamaguchi N.;
 RT "Cloning and characterization of a novel serine protease, mbssp-4.";
 RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; AB010779; BAB20263.1; -

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: AB010778; BAB2062.1; -
 CC InterPro: IPR001314; Chymotrypsin.
 CC InterPro: IPR001254; Trypsin.
 CC Pfam: PF00089; trypsin.1
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00020; TRYP_SPEC.1.
 CC PROSITE: PS50240; TRYPSIN_DOM.1.
 CC PROSITE: PS00134; TRYPSIN_HIS.1.
 CC PROSITE: PS00135; TRYPSIN_SER.1.
 CC Hydrolase: Serine protease; Signal.
 CC SIGNAL
 CC CHAIN 1 32
 CC ACT_SITE 33 306
 CC ACT_SITE 90 90
 CC ACT_SITE 141 141
 CC ACT_SITE 242 242
 CC ACT_SITE 75 91
 CC ACT_SITE 175 248
 CC ACT_SITE 208 227
 CC ACT_SITE 238 266
 CC ACT_SITE 70 70
 CC CARBOHYD
 CC SEQUENCE 306 AA; 33262 MW; FBBF03C0C285E7E8 CRC64; (POTENTIAL).

Query Match 44.38; Score 696; DB 1; Length 306;
 Best Local Similarity 47.68; Pred. No. 4.9e-36;
 Matches 136; Conservative 47; Mismatches 87; Indels 16; Gaps 6;

QY 6 AVPLLLLCFSSORAKATA-----CGRPMILNRMVSGQDQOEGEMPVOVSITORNGSHFC 60
 DB 16 SLILLVLTSTAPISATINIVSPCGKQDLNRTVGSEDMQMPVIVSLKNGSHHC 75
 QY 61 GGSILAEQWVLTAAHCFR-NTSETSLYVLLGAROLVOPGPHAMARVQYESNPLY--- 116
 DB 76 AGSLTNMNVVTAACHFKSNMDKPSLFEVLGAMKLGSPGSPKVGIAVWLPHPRYSWK 135
 QY 117 OCTASSADVAVELAPVPTNYILPVCLPSPVFEFGNMCMQWTGMSPSSEEDLPPPR 176
 DB 136 EST--HADIALVRLHSHSIOFSERILPCLPSSSVALLPKTKDWMAGWSTIDGVPPLPHQ 193
 QY 177 ILQKLVAVIIPDKCNLLYSKDTFEGYQPTKINDMLCAGEEGKKDCKDGGGPIVCL 236
 DB 194 TLQKLVKPIIDSELCSLYWR---GAQGEAITTEGMLCAGYLEGRDACLGGSGGLMCQ 249
 QY 237 VQOSMLQAGVSWGEGCARQNPQVYIRVTAHNNHIIKPLQDQ 282
 DB 250 VDDHMLTGLIISWEGCA-DDRPGVYTSLLAHRSVQRIYOGVQV 294

RESULT 4
 PSS8_HUMAN STANDARD; PRT; 343 AA.
 AC Q16651; Q9UCA3;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Prostatein precursor (EC 3.4.21.-).
 GN PRS58.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=prostate;
 RX MEDLINE=95286644; PubMed=7768952;

RA Yu J.X., Chao L., Chao J.;
 RT "Molecular cloning, tissue-specific expression, and cellular
 RT localization of human Prostatein mRNA."
 RL J. Biol. Chem. 270:13483-13489(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 45-64.
 RX MEDLINE=94308140; PubMed=8034638;
 RA Yu J.X., Chao L., Chao J.;
 RT "Prostatein is a novel human serine proteinase from seminal fluid.
 RT Purification, tissue distribution, and localization in prostate
 RT gland."
 RL J. Biol. Chem. 269:18843-18848(1994).
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A
 CC DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF
 CC ITS C-TERMINUS.
 CC -1- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,
 CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR
 CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL
 CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL
 CC FLUID.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: LA1351; AAC41759.1; -
 DR EMBL: U33446; AAB19071.1; -
 DR EMBL: BC001462; AAB01462.1; -
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.159; -.
 DR MIM: 600823; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPEC.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS.1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 DR Hydrolase: Serine protease; Zymogen; Signal; Glycoprotein;
 KW Transmembrane.
 FT SIGNAL 1 29
 FT PROPEP 30 32
 FT CHAIN 33 44
 FT CHAIN 45 322
 FT PROPEP 323 343
 FT TRANSMEM 320 340
 FT DOMAIN 45 286
 FT DISULFID 37 154
 FT DISULFID 70 86
 FT DISULFID 168 244
 FT DISULFID 201 223
 FT DISULFID 234 262
 FT ACT_SITE 85 85
 FT ACT_SITE 134 134
 FT ACT_SITE 238 238
 FT CARBOHYD 159 159
 FT SEQUENCE 343 AA; 36431 MW; 98D06447F5A8C1B2 CRC64;

| Query Match | Best Local Similarity | Score 614.5; DB 1; | Length 342; |
|---------------------------|-----------------------|--------------------|-------------------|
| Matches 119; Conservative | 42.5%; | Pred. No. 1.4e-48; | Indels 13; Gaps 4 |

[illegible]

| RESULT | 6 |
|------------|---|
| TRYG_HUMAN | |
| ID | TRYG_HUMAN |
| AC | Q9NR08: Q9NR08; Q9UBR2; |
| CC | Q9NR08: Q9NR08 (Rel. 40); Created; |
| DT | 16-OCT-2001 (Rel. 40); Last sequence update) |
| DT | 16-OCT-2001 (Rel. 40); Last annotation update) |
| DE | tryptase gamma precursor (EC 3.4.21.-) (Transmembrane tryptase). |
| GN | TPSG1 OR TMT. |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| OX | NCBI_Taxid-9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2). |
| RP | MEDLINE-20302813; Pubmed-10843716; |
| RA | Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaoiro M., |
| RA | Wolters P.J., Verghese G.M.; |
| RT | "Characterization of human gamma-tryptases, novel members of the |
| RT | chromosome 16p mast cell tryptase and prostatic gene families."; |
| RL | J. Immunol. 164:5566-5575(2000). |
| RN | [2] |
| RP | SEQUENCE FROM N.A. |
| RP | MEDLINE-99452974; Pubmed-10521469; |
| RA | Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C., |
| RA | Friend D.S., Krilis S.A., Stevens R.L.; |

| ID | TRYG_MOUSE | STANDARD | PRT | 311 AA |
|------------|--|---|-----|-----------------------------|
| Db | 9 | LLLAIVEVSLRTLQPGCGRPQVSDAGRIYVGHNAAPAMPQWQASLRLRRHHVCGSLL | 68 | |
| Qy | 66 | AEQVYLTAHCFRNTSETSLYQVLLGANOYQPGPHMYARVROY--ESNPLYQGTASSA | 123 | |
| Db | 69 | SPQWLTAHCFHSGSLNSDYYHGLGELEITL-SPH--FSTVRLQILHSSPSGQ--PRTSG | 124 | |
| Qy | 124 | DVALVELAPVPFNWYLLVPCLPDPSVLEFDGMNQMVMGKSPSEDDLLPEPRILQKAV | 183 | |
| Db | 125 | DIALVELSVPTLSSRLIPVCLPEASDQFCGIRCMVWKGCGYTRGEGPLPPYSIREKV | 184 | |
| Qy | 184 | PLIDPKCNLLYSKQTEFGYQPKTKNMLDAGEFEGRACKDGSGSPLYCLVQGSWLQ | 243 | |
| Db | 185 | SVVDERTCRROY-----GPGSILQPMCLA---RGGDACQDDSGSPLYCQVNGAVQ | 236 | |
| Qy | 244 | AGVISWEGECARQNRPGVYIRVTAHHNMIHRI | 276 | |
| Db | 237 | AGIVSWEGECGRPNRPVYTRVPAVNMIRRH | 269 | |
| RESULT | 7 | | | |
| TRYG_MOUSE | TRYG_MOUSE | STANDARD | PRT | 311 AA |
| AC | 09OUL7 | | | |
| Df | 16-OCT-2001 | (Rel. 40, Created) | | |
| Df | 16-OCT-2001 | (Rel. 40, Last sequence update) | | |
| Df | 16-OCT-2001 | (Rel. 40, Last annotation update) | | |
| Df | 16-OCT-2001 | (Rel. 40, Last annotation update) | | |
| De | Trypase gamma precursor (EC 3.4.21.-) (Transmembrane tryptase). | | | |
| GN | TPSG1 OR TMT. | | | |
| OS | Mus musculus (Mouse). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| OX | NCBI_Taxid:10090; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM NP.A. | | | |
| RC | STRAIN-129SV, AND BALB/C; | | | |
| RC | MEDLINE:99452974; PubMed:10521469; | | | |
| RX | Wong G-W., Tang Y., Feyfant E., Sall A., Li L., Li Y., Huang C., | | | |
| RA | Friend D.S., Killis S.A., Stevens R.L.; | | | |
| RT | "Identification of a new member of the tryptase family of mouse and | | | |
| RT | human mast cell proteases which possesses a novel COOH-terminal | | | |
| RT | hydrophobic extension."; | | | |
| RL | J. Biol. Chem. 274:30784-30793(1999). | | | |
| CC | -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (potential). | | | |
| CC | -1- TISSUE SPECIFICITY: Expressed in many tissues. | | | |
| CC | -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | | |
| CC | TRYPSIN FAMILY. TRYPTASE SUBFAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way that | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-slb.ch/announce/ | | | |
| CC | or send an email to license@isb-slb.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; AF115760; AAF03698.1; - | | | |
| DR | EMBL; AF115523; AAF03696.1; - | | | |
| DR | HSSP; P20231; IAAO. | | | |
| DR | MED; MGI:1348391; Tpsgl. | | | |
| DR | InterPro; IPR001314; Chymotrypsin. | | | |
| DR | InterPro; IPR001254; Trypsin. | | | |
| DR | Pfam; PF00089; trypsin; 1. | | | |
| DR | PRINTS; PR00782; CHYMOTRYPSIN. | | | |
| DR | SMART; SMO0020; Tryp_SPC; 1. | | | |
| DR | PROSITE; PS00240; TRYPSIN_DOM; 1. | | | |
| DR | PROSITE; PS00134; TRYPSIN_HIS; 1. | | | |
| DR | PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG. | | | |
| DR | Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen; | | | |
| KW | Transmembrane. | | | |
| FT | SIGNAL | 1 | 16 | POTENTIAL. |
| FT | CHAIN | 17 | 28 | TRYPTASE GAMMA LIGHT CHAIN. |
| FT | CHAIN | 30 | 311 | TRYPTASE GAMMA HEAVY CHAIN. |
| FT | TRANSMEM | 277 | 297 | POTENTIAL. |

| | | |
|----------------|------------------|-----------------------------------|
| EMBL; D31789; | BAA06598..1 | - |
| HSSP; P20231. | IAAO. | |
| DR MEROPS; | S01.143; | - |
| DR InterPro; | IPR001314; | Chymotrypsin. |
| DR InterPro; | IPR001254; | Trypstin. |
| DR pfam; | PF00089; | Lypsin; 1. |
| DR PRINTS; | PR00722; | CHYMOTRYPSIN. |
| DR SMART; | SM00020; | TRYP_SPE; 1. |
| DR PROSITE; | PS02040; | TRYPSIN_DOM; 1. |
| DR PROSITE; | PS00134; | TRYPSIN_HIS; 1. |
| DR PROSITE; | PS00135; | TRYPSIN_SER; 1. |
| KM Hydrolyase; | serine protease. | Signal; Glycoprotein. |
| FT SIGNAL | 1 | POTENTIAL. |
| FT CHAIN | 26 | 270 |
| FT ACT_SITE | 69 | 69 |
| FT ACT_SITE | 116 | 116 |
| FT ACT_SITE | 219 | 219 |
| FT DISULFID | 54 | 70 |
| FT DISULFID | 150 | 225 |
| FT DISULFID | 183 | 206 |
| FT DISULFID | 215 | 243 |
| FT CARBOHYD | 46 | 46 |
| FT CARBOHYD | 127 | 127 |
| FT CARBOHYD | 228 | 228 |
| SO SEQUENCE | 270 AA; | 30166 MW; 1BEI02DB86943401 CRC64; |

| | | | |
|----|-----|--|-----|
| QY | 6 | APVILLILLCFSGQRRKKAATACGPRMLNRVYAGQDQDEGEWVQSIQKNGS--HFCGG | 62 |
| | | : : : : : : : : : | |
| Db | 8 | ALPFLSLM---HRSPLQCEWG-----IVGQDEAPGKKMPOVSLRANETWYRHFCGG | 56 |
| QY | 63 | SLIAEQWVLTAAHCERNT-SETSLYQVILGARLVQPRPHAYARARQVESPLVQCGAS | 122 |
| | | : : : : : : : | |
| Db | 57 | SLIHPRQVLTAAHCGRPIADPNKVRVQGLKQVLY---HDHLAVSRKLIITHPTVATON | 113 |
| QY | 122 | SNADVAVLELEAPVPTNTILPVCLPDESVYIETGMNCWNTGMSPEEDLILPEPRILQKL | 181 |
| | | : : : : : : : : : | |
| Db | 114 | GDVIALLEKELKNNVNTSSIHVHVS.LPRASETFPSCGLCAWNTGMDINDVSLPEPRILEKV | 173 |
| QY | 182 | APVILIDTPKCNLYLSKDEFEFGYQPKTATKNMLCAGEEGRKQACGDSGGPLVCLVQSGS | 24 |
| | | : : : : : : : : : | |
| Db | 174 | QPVVVENQGLDCLKKHYKGYTSDNLIHYRDMCLAG--NEG-HHSCGQDSGGLPVCKVNGTW | 233 |
| QY | 242 | LAQGYISNGEGCAGAKONRPGVYIRTATAHNNHTIIRK | 278 |
| | | : : : : : | |
| Db | 232 | LAQGYVSWEGCAGALPRRGILITRYATYLLIMIHRYPR | 268 |

| | | | | |
|-------|--|-----------|------|---------|
| ID | MCT7_RAT | STANDARD; | PRT: | 273 AA. |
| AC | P27435; P27436; | | | |
| DT | 01-AUG-1992 (Rel. 23, Created) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | |
| DE | Mat mast cell protease 7 precursor (EC 3.4.21.59) (RMCP-7) (Trypsin, | | | |
| DE | skin). | | | |
| GN | MCP7 OR MCP7. | | | |
| OS | Rattus norvegicus (Rat). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | |
| OX | NCBI_Taxid=10116; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=SPRAGUE-DAWLEY; | | | |
| RX | MEDLINE=9719430; PubMed=8996238; | | | |
| RA | Lutzel-Schwab C., Pejler G., Aveskogh M., Hellman L.; | | | |
| RT | "secretory granule proteases in rat mast cells. Cloning of 10 | | | |
| RT | different serine proteases and a carboxypeptidase A from various rat | | | |
| RL | mast cell populations." | | | |
| RL | J. Exp. Med. 185:13-29(1997). | | | |
| RN | [2] | | | |
| RP | SEQUENCE OF 29-53. | | | |
| RC | STRAIN=SPRAGUE-DAWLEY; TISSUE=SKIN; | | | |
| RX | MEDLINE=91242400; PubMed=2036367; | | | |
| RA | Braganza V.J., Simmons W.H.; | | | |
| RT | "tryptase from rat skin: purification and properties."; | | | |
| RL | Biochemistry 30:4997-5007(1991). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 29-51. | | | |
| RC | TISSUE=Breast carcinoma; | | | |
| RX | MEDLINE=92231826; PubMed=1314562; | | | |
| RA | Eco I., Grubbs C.J.; | | | |
| RT | "separation, purification and N-terminal sequence analysis of a novel | | | |
| RT | lupercin-sensitive serine endopeptidase present in chemically | | | |
| RL | induced rat mammary tumour." | | | |
| BL | Biochem. J. 283:209-216(1992). | | | |
| -I- | FUNCTION: TRYPSIN IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST | | | |
| CC | CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION | | | |
| CC | RESPONSE OF THIS CELL TYPE. | | | |
| CC | -I- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-, Lys-I-, but | | | |
| CC | with more restricted specificity than trypsin. | | | |
| CC | -I- SUBUNIT: HOMOTETRAMER. | | | |
| CC | -I- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON | | | |
| CC | MAST CELL ACTIVATION. | | | |
| CC | -I- TISSUE SPECIFICITY: MAST CELLS. | | | |
| CC | -I- PTM: GLYCOSYLATED (PROBABLY). | | | |
| CC | -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE | | | |
| CC | TRYPSIN FAMILY. TRYPSIN SUBFAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| EMBL: | U67910; AABA8263.1; -. | | | |
| DR | PIR; A23698; A23698. | | | |
| DR | PIR; S21275; S21275. | | | |
| DR | HSSP; P20231; IAAO. | | | |
| DR | MEROPS; S01_026; -. | | | |
| DR | InterPro; IPRO01314; Chymotrypsin. | | | |
| DR | InterPro; IPRO01254; Trypsin. | | | |
| DR | Pfam; PF00089; trypsin; 1. | | | |
| DR | PRINTS; PR00722; CHYMOTRYPSIN. | | | |
| DR | SMART; SM00020; Tryp_Spc; 1. | | | |
| DR | PROSITE; PS50240; TRYP SIN_DOM; 1. | | | |
| DR | PROSITE; PS00134; TRYP SIN_HIS; 1. | | | |
| DR | PROSITE; PS00135; TRYP SIN_SER; 1. | | | |
| DR | MultiClass; Serine protease; Glycoprotein; zymogen; signal; | | | |
| DR | MultiGene family | | | |
| DR | MultiGene family | | | |

| FT | SIGNAL | 1 | 18 | POTENTIAL. |
|----|----------|---------|-----------|--------------------------------------|
| FT | PROPEP | 19 | 28 | ACTIVATION PEPTIDE. |
| FT | CHAIN | 29 | 273 | MAST CELL PROTEINASE 7. |
| FT | ACT_SITE | 72 | 72 | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| FT | ACT_SITE | 119 | 119 | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| FT | ACT_SITE | 222 | 222 | CHARGE RELAY SYSTEM (BY SIMILARITY). |
| FT | DISULFID | 57 | 73 | BY SIMILARITY. |
| FT | DISULFID | 153 | 228 | BY SIMILARITY. |
| FT | DISULFID | 186 | 209 | BY SIMILARITY. |
| FT | DISULFID | 218 | 246 | BY SIMILARITY. |
| FT | CARBOHYD | 49 | 49 | N-LINKED (GLCNAC. . .) (PROBABLE). |
| FT | CONFLICT | 42 | 42 | W -> V (IN REF. 3). |
| FT | CONFLICT | 49 | 51 | NDP -> WLP (IN REF. 3). |
| SO | SEQUENCE | 273 AA; | 30400 MW; | 65A5EDAD279FB284 CRC64; |

Query Match 37.6%; Score 590.5; DB 1; Length 273;
Best Local Similarity 45.6%; Pred. No. 1.7e-46;
Matches 125; Conservative 43; Mismatches 95; Indels 11; Gaps 6;

| | | |
|----|---|-----|
| 9 | LLLLLCSORAKATACGPRRLNRMVGGQDQEGEMWQVSIQRNGS---HFCGSLI | 65 |
| 5 | LLLTPLLSLVHAAPSLAMP--EGIVGGQELASGNKMPQVSLRVNDTYMHHFCGSLI | 62 |
| QY | AEQWVLTAAHCF-RMTSETSLYOVLGARQLVOPGPHAMYARVROVESNPLYOGTASSAD | 124 |
| DB | HPQWVLTAAHCF-RMTSETSLYOVLGARQLVOPGPHAMYARVROVESNPLYOGTASSAD | 119 |
| QY | VALVELAPVPTNYIILVCLPDPSTVEFTGMNCWYTGSGPSEEDLLPEPRLIOLAMP | 184 |
| DB | IALLKLTNPVNTISNHTVSLPPASETFPSGTLCWYTGSGPSEEDLLPEPRLIOLAMP | 179 |
| QY | IIDPRKCNLYSKDTERGYQPKTIKNDMLCAFEEGKKDKGSGGPIVCLVGSQSLQA | 244 |
| DB | IENRRLCDLKYHKGNTGDNVNHVTRDMLCAG-NEG-HDSCGGSGGPIVCLVGSQSLQA | 237 |
| QY | GVISMGECAQONRPGYIRVTAHNMVIRIRIPK | 278 |
| DB | GVISMGECAQONRPGYIRVTAHNMVIRIRIPK | 271 |

Search completed: August 13, 2002, 09:00:50
Job time: 227 sec

us-10-041-006-7.p2n.rge

| | | | | | | | | |
|-----------------|---|--------|--------|---------|------|--|----------|-----------------------|
| gb_pat:AR080457 | + | 594.00 | 931.85 | 1.5e-43 | 1219 | | AR080457 | Sequence 7 from pa |
| gb_to:NMIMC1 | + | 594.00 | 931.85 | 1.5e-43 | 1219 | | D31789 | Mongolian gerbil MNIN |
| gb_ro:RW67910 | + | 590.50 | 926.15 | 3.2e-43 | 1215 | | U67910 | Rattus norvegicus man |
| gb_pat:AR080461 | + | 588.50 | 922.79 | 4.3e-43 | 1128 | | AR080461 | Sequence 15 from p |
| gb_pt:HMTRY2A | + | 588.50 | 923.79 | 4.3e-43 | 1128 | | M34392 | Human tryptase-II mR |

[illegible]

| LOCUS | DEFINITION | VERSION | KEYWORDS | SOURCE | ORGANISM | REFERENCE | AUTHORS |
|-----------|--|------------|-------------|---------------------|---|-----------|---------|
| HSA306593 | Homo sapiens mRNA for marapsin (MPN gene). | AJ306593.1 | GI:13516325 | marapsin; MPN gene. | Homo sapiens | 1013 bp | linear |
| | | | | | human. | | |
| | | | | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| | | | | | 1 (bases 1 to 1013) | | |
| | | | | | Fortunato,M., Dando,P.M., Rawlings,N.D. and Barrett,A.J. | | |

protease

REFERENCE 2 /bases 1 to 10131

AUTHORS
Fortunato, M.

Laboratory, Babraham Institute, Babraham Hall, Babraham, Cambridge

Source

```

/gene="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE 2698030"
20..892
/gene="MPN"
20..892
/gene="Mpn"
/function="putative trypsin-like activity"
/codon_start=1
/product="marapsin"
/protein_id="CAC35467.1"
/db_xref="GI:13516326"
/translation="MRPAVALPLLLICGSSORAKATAAGRPRLNRWVGODPDQEG
EMWVYSIQNSHQSGSLIAQWLTAAHCRRNTSERISLYVLGAQLVDPGFHA
MYARKOVESNPPLYQSTASADVALELAEVPFTNYILPVCILPDPSVLFETGMNV
TGKSSSEEDLEPERILLQKLAVPIIDTPRCNLVSKDEFGYOPTIKNDMLCAFE
EKGDKDCKDGSGGPLVCLVGOEWLAGVISMEGCARONRPGVIIVTATHNHHRIRI
PKLOFPARGCGKK"
BASE COUNT      192 a       334 c       321 g       166 t
ORIGIN

alignment_scores:
Quality: 1571.00          Length: 290
Ratio: 5.417              Gaps: 0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-10-041-006-7 x HSA306593 ..

Align seg 1/1   to: HSA306593   from: 1   to: 1013

1 MetargatgProalaaAlavAlProleuleuleuleuleucysphnglyse 17
|||||
20 ATGAGGCGGCCGGCGGCTCCGCTCCTGCTGCTGTATTGGGCTC 69

17 rglmargAlaylsAlaatPhralacysglyArgproargmetleusna 34
|||||
70 TCAAGAGGGCCAAAGCACACAGCCTGTGTGGCCCCAGAGATGTGAACC 119

34 rgmetvalglylglnaspthrcingnuglgyltutptprctrglnaval 50
```


[illegible]

| | | | |
|--------------------------|---|---------|--------|
| seq_documentation_block: | | | |
| LOCUS | E31782 | 1109 bp | DNA |
| DEFINITION | Novel compound. | | Linear |
| ACCESSION | E31782 | | |
| VERSION | E31782.1 GI:13025811 | | |
| KEYWORDS | JP 1999098992-A/1. | | |
| SOURCE | unidentified. | | |
| ORGANISM | unidentified. | | |
| REFERENCE | 1 (bases 1 to 1109) | | |
| AUTHORS | Nicola, A. B. H. H. and Kurinkenhlaado, C. D. S. S. | | |
| TITLE | Novel compound | | |
| JOURNAL | Patent: JP 1999098992-A 1 13-APR-1999; | | |
| COMMENT | SMITHKLINE BEECHAM CORP PUBLIC LTD CO | | |
| | OS unidentified | | |

```

PR 05-JUN-1998 JP 1998199425
PF 10-JUN-1998 GB 97120808.5,17-OCT-1997 GB 97308295-1, PR
20-FEB-1998 GB 9803650.2
PI NICOLA ANN BAUESU, HELEN ELIZABETH KUBINKENBIAOD, PI
CHRISTOPHER DONALD SAZAN
PC C12N5/09,A6IK31/7/0,A6IK35/7/6,A6IK38/0/0,A6IK38/0/0
PC A6IK39/395.
PC A6IK39/395,A6IK48/0/0,C07K16/4/0,C12N5/1/0,C12P21/02,C12P21/08
PC C12Q1/68,
PC C12N5/00,A6IK37/02,A6IK37/02,A6IK37/02,C12N5/00 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..1109
ET 'organism=' 'Unidentified'.

```

| FEATURES | Location/Qualifiers |
|------------|--------------------------|
| source | 1..1109 |
| | /organism="unidentified" |
| | /db_xref="taxon:32644" |
| BASE COUNT | 229 a 344 c 344 g 192 t |
| ORIGIN | |

[illegible]

```
alignment_block:
US-10-041-006-7 x E31782 .
```

Align seg 1/1 to: E31782 from: 1 to: 1109

[illegible][illegible]

[illegible]

| | | | |
|------------|--|--|----------------------------|
| | 648 | ATCTTTGAGACGGGCATGAACCTCCTGGGTACTCGTGGCGGCCGCCAC | 697 |
| | 167 | rGlUGluAspLeuProGluInProKarglleleGlnIlySleAlaValP | 184 |
| | 698 | TGAGGAGAACCTTCCTCCCGAACCCGGGATCCGCAAGAACTCGTGTCG | 747 |
| | 184 | rollelleasprThrProlysCysAsnLeuLeuYrSerIySaspThrlu | 200 |
| | 748 | CcATCATCGACACACCAGSTGCACCTGCTCACGACAAGACACGGAG | 797 |
| | 201 | PheGlyTYrGlnProIystrhrIlelySasnspmetleuCysAlacly.P | 217 |
| | 798 | TTTTGGCTACCAACCCAAAACATCAAGAATGACATCTGTGCCCGGNTT | 847 |
| | 217 | heeIUGluGlyLySLySaspAlaCysLys.GlAYSPserGlyIypProe | 233 |
| | 848 | TCAGAGGGGCGAAGAGATGCTCGAAGTGGG.....GGAGCCAT | 888 |
| | 233 | vUaLCysLeuValGlyGlnserTrpLeuGlnalaclyVal...lleSeXT | 249 |
| | 889 | GtGGGTCTTCATGGGATTCAGTTCACCGNACGGCCAGGCCCAAGGGT | 938 |
| | 249 | rpolyluglucly.....CysAlarglGlnsnagProgly | 260 |
| | 939 | GGGGTTAAGGCGCTGTCCATGGGTTATGCTTCAACAGAGANCGCCAGT | 988 |
| | 261 | VaLYrrIleArgValThrAlaHisHisAsnTrpIleHisArgIlelePr | 277 |
| | 989 | GtCTAATATCGGTGTNAACGGCCAAACAACTGATCTCATCGATCATNC | 1038 |
| | 277 | OlySLeuGlnPheGlnProAlaArgleugly | 287 |
| | 1039 | CAAACCTGAG.TTCCAGCCCAACGAGATTGTGC | 1068 |
| | seq_name: | gb_pat:E31784 | |
| | seq_documentation_block: | | |
| LOCUS | E31784 | 1109 bp | DNA linear PAT 07-FEB-2001 |
| DEFINITION | Novel compound. | | |
| ACCESSION | E31784 | | |
| VERSION | E31784.1 GI:13026588 | | |
| KEYWORDS | JP 1999098992-A/3. | | |
| SOURCE | unidentified. | | |
| ORGANISM | unclassified. | | |
| REFERENCE | 1 (bases 1 to 1109) | | |
| AUTHORS | Nicola,A.B.H.H. and Kurinkenbiado,C.D.S.S. | | |
| TITLE | Novel compound | | |
| JOURNAL | Patent: JP 199098992-A 3 13-APR-1999; SMITHKLINE BEECHAM CORP PUBLIC LND CO | | |
| COMMENT | OS Unidentified PN JP 1999098992-A/3 PD 13-APR-1999 PE 09-JUN-1998 JP 1998199425 PF 10-JUN-1997 GB 9712088.5,17-OCT-1997 GB 97308295-1, PR 20-FEB-1998 GB 9803650.2 PI NICOLA DONALD BAJESU HELEN ELIZABETH KURINKENBIADO, PI CHRISTOPHER ANN DALL SAZAN PC C12N15/09,A61K31/70,A61K35/76,A61K38/00,A61K38/00,A61K38/00, PC A61K39/395, PC A61K39/395,A61K48/00,C07K16/40,C12N5/10,C12P21/02,C12P21/08, PC C12O1/68, PC C12N15/00,A61K37/02,A61K37/02,A61K37/02,C12N5/00 CC Strandedness: Single; Topology: Linear; CC Topology: Linear; FH Key Location/Qualifiers FT source 1..1109 /organism='Unidentified'. FT location/Qualifiers 1..1109 /organism='unidentified' /db_xref='taxon:32644' | | |
| FEATURES | | | |
| source | | | |
| BASE COUNT | 224 a 327 c 338 g 191 t 29 others | | |


```

DEFINITION Homo sapiens chromosome 16 clone CTD-2593F12, complete sequence.
AC093517
VERSION AC093517.2 GI:15808554
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 160601)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
REFERENCE 2 (bases 1 to 160601)
AUTHORS Unpublished
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE 3 (bases 1 to 160601)
AUTHORS Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL DOE Joint Genome Institute.
REFERENCE 3 (bases 1 to 160601)
AUTHORS Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
JOURNAL On SEP 29, 2001 this sequence version replaced gi:15383806.
COMMENT
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source
location/Qualifiers
1..160601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2593F12"
BASE COUNT 37601 a 43383 c 43006 g 36611 t
ORIGIN
1
22 AlaAaThaAlaAcYsGlyArGPrOaGmCtLeuAaASrArGmValGlyG 38
111
100378 GCTCCCCCAAGCCTGTGTGTCGCCCAAGAGTGTGAACCGAATGTGTGGCGG 1003329
38 yGlnAsPThnGlnGluGlyGlnTrpProTrpGlnValSerIleGlnArgA 55
GCAGGACACCCAGAGAGGGCGAGTGGCCCTGGCAAGTCAGCATCCAGCGCA 100279
55 snGlySerHisPheCySGlyGlySerLeuIleAlaGlnGlnTrpValIleu 71
111

```

100278 ACGGAGCCACTTCTGCGGGGCGAGCCTCATCGGAGCAGTGGTCTTG 100229
72 ThrAlaAlaHisCysPhe..... 77
100228 ACGGCTGCGACTGCTTCCGAAGTAGTCCGCCGCCCTGCCCCGCC 100179
77 77
100178 CATAGCCGTGACAGCGCCCGCGCGCAGCGGTTCAGCACCGTGGACAGC 100129
77 77
100128 GCCCGCGCGCAAAATCTGGGGGTGACTTCCTGGGGCTCTGTGCA 100079
77 77
100078 GCCCCTCCACCCAGATCTTCCCTTAGTCCAACTCCAGGGCTAATCTC 100029
77 77
1028 CAGTTGCAACGGCTGCTCCCGCCCGGGAGGTGCTCGCACCGCCCCC 99979
77 77
99978 GACCCCTCATCCCTTCACACCACCTACCCACTCCTGTGGGTCCCTGC 99929
77 77
99928 AGAAGCGCGCGAGGCTGTCGCCACCGCCCTCTGCGCTTTCGCCA 99879
77 77
99878 TCCCGCACACACTCAGCTCCAGGACACTTTCGGGAGAACTGCT 99829
77 77
99828 CACAAAGCCCAAGACAGACAGAAAGCCCTTCCTCCCTACCCACT 99779
77 77
99778 GAACCACCCAAGAAAGCCCTGAGCAGAGCCAGCCACCCCTCTGC 99729
77 77
99728 CATGATGAACACCTGCTGCCACACTTCGGGGTTCGCCAGGCCCTC 99679
77 77
9678 ACCTCACACTCAACACCGCAGCTCTAATATTTTAAACCCACATCTTT 99629
77 77
99628 TTCTTTTCTCTCTGATCTTTAAAGAAATATCATGACAAAAAAC 99579
77 77
99578 CCACATCTTAATTCAGATCTCAAGCGGCAGCAGGTGGCTCAACCCG 99529
77 77
99528 TAATCCAGCACTTGGAGGCGCAAGCGGCAGATCAGTTGAGCCAGG 99479
77 77
99478 AGTTCAAGACCAAGCCGGGCAACACAGACACCTGTCTACTAAAAA 99429
77 77
99428 TAAAAAAATTTAAAAACAAAAACAACAAAAAACCCAGGTG 99379
77 77
99378 CAGTGTGCTGTGTGTCACCAAGTACTCAGAAAGTGAAGCGAGAGATGG 99329

77 77
99328 CTTGAGCCAGAGTTGGAGGCTGCAGTAGCTACAGTTGACTCTGCA 99279
77 77
99278 CTCACGCTGGGTGACAGAAATGAGAACTCTGTCTAGGGGAAAAAAA 99229
77 77
99228 ATCGAACTCGGGCTGAGAGAGGCGAGTCACTGCCAGAGCCCGCAG 99179
77 77
99178 GTGGGGCTGGCCCTGCTGCACCTGGAGGGTGGTTCTCTCCGAGCCAGCT 99129
77 77
99128 TTCAGTAGGCACTGTACTTCCACACAGACAGGTGAGAAAGCTGAT 99079
77 77
99078 CACAGAAACTTGTTAAGTACTTGCACCAAGGCCACACAGCGGCCAGGG 99029
77 77
99028 TGGAGTGGAAATCTGATGTCGGGGGCTGCGCTTCAATGTCAACAGCAT 98979
77 77
98978 CACCGCTGCACCTGCTGCTCTAGAGATGACTCTCGAAGACATTTGACAG 98929
77 77
98928 AGAGAGGTGTGGGCAAGGTGTGCCAAGAGACCCAGCCCTCCCG 98879
78 Ar 78
98878 GCCTGCCCCACACAGAGGGCCGCTCAGCGCCCTCTGTTTCGCG 98829
78 Ar 78
98828 CACAGCACTCTGAGAGCTCCCTGACAGGTCTGCTGGGGGCAAGGAG 98779
78 g.AsnThrSerGluThrSerLeuTyrGlnValLeuLeuGlnValArgGln 94
98778 CTAGTGCAGCGGGGACACACGCTATGTATGCCGGGTGAGGAGGTGA 98729
111 userAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeuV 128
98728 GAGCAACCCCTGCTACAGGCGACGGCTCCACGCTGACGTGGCTGG 98679
128 ALGluLeuGlnAlaProValProPheThrAsnTyrIleLeuProValCys 144
98678 TGGAGCTGGAGGACCAAGTGCCTTACCAATTAATCCCTCCCGTGTG 98629
145 LeuProAspProSerValIlePheGluThrGlyMetAsnCysTrpValIb 161
98628 CTGCTTACCCCTGCTGATCTTTGAGACGGGCATGACTGCTGGTCA 98579
161 rGlyTrpGlySerProSerGlu..... 169
98578 TGGCTGGGCGAGCCCGCAGTGAAGGTAAAGGAGACAGGCTGGAAAG 98529
169 169
98528 ATGGGGATATGCTCAAGAAGGCCAGAGCGCTTGGGACCACTGCA 98479
169 169
98478 ACCCAATCTCTCTGTGCTAATTAATATACATACACAGGCTCTGTT 98429


```

169 ..... 169
98428 ACACAGGTGGGCGGAGGAGGAGCAGACAGACAGAGAGAGAGAGG 98379
169 ..... 169
98378 TGAAGGAGAGGGGAGACAGAAAGGCCAGCTGCTGCTGAAGAGAGG 98329
169 ..... 169
98328 AGAGTGTCTGCCCCACAGTCCCTTAAGCCAGAGATGTCTAAACATGAG 98279
169 ..... 169
98278 AAACGACGAGCAGAGATGAGTGTGTTCGGGGCCGCTGGTCTTGAT 98229
170 ..... AspLeu 172
98228 GAGGAGTCCGTTGAGCCCTGGCTGTCCGCCCTCTCCAGACCTCTG 98179
173 ProGluProArgIleLeuGlnLysLeuAlaValProIleLeuSphPr 189
98178 CCGGAGCCGGGATCCGCGAGAAACCTGCTGCTCCATCATCACACACC 98129
189 OlyCysAsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProL 206
98128 CAAGTGCACCTCTCTACAGCAAGACACCGAGTTGGCTACCAACCA 98079
206 YstHrIleLysAsnAspMetLeuCysAlaGlyPheGluGlyLysLys 222
98078 AAACCATCAAGATGACATGCTGTGCGCGGCTCGAGGAGGAGAGAG 98029
223 AspAlaCysLys ..... 226
98028 GATGCTGCAGAGGTGGAGCAATGTGTCTCATGGGACTGATCCCA 97979
226 ..... 226
97978 CCGCCAGGCCAGGCGAGGGTGGATATGCCCCGCTCATGGGCTATGC 97929
226 ..... 226
97928 ATTCACACAGCTGAGCAGCTGTGTTTGAAGAGCAGATTCCAGGCGGG 97879
226 ..... 226
97878 TGTGTGTCTCAGCGCTTAATCCAGCACTTGGAGGCCAAGAGGCT 97829
226 ..... 226
28 GGATCACAAGTCAAGAGATGAGAGCCATCTGGCTAACACAGTGAACC 97779
226 ..... 226
97778 CCATCTACTATAAATACAAAAATTAGCCAGGCGTGTGTGGGCGCC 97729
226 ..... 226
97728 TGTAGTCCAGCTACTTGGAGGCTGAGCAGAGATGGCGCAACCTG 97679
226 ..... 226
97678 GGAAGGCGGAGCTTGCAGTGAGCCGAGATGCCGCACTCCAGCCTG 97629
226 ..... 226
97628 GGCACAGAGCAAGACTCCGTCTCAATAAAAAAAGAAAAAGACAGA 97579
226 ..... 226
97578 TTCCTGGACCCCAACCCAGAGATCAGATGATGCTGGGGTGGGG 97529
226 ..... 226

```

```

97528 CACAGAAACTGATTTTCAACACACACACACACACACACACACACAC 97479
226 ..... 226
97478 ACACCTAAGAAATAGCAGCTTAGACATCGGACTGGAGGGAGTGGAG 97429
226 ..... 226
97428 CCGGCGAGAGAGCTCTGGGATGGAGAGAGAAATCCATTTCGAGACTAC 97379
226 ..... 226
97378 ACACCTCTGTGCGGCTGCAACTGAGTCTTGGGAATCAAGTGGGCTC 97329
227 ..... GlyAspSerGlyGlyProLeuVal 234
97328 TGACGGTACTCTGCGTCCCGGAGGGGAGCTCGGGGCGCCCTGGTG 97279
235 CysLeuValGlyGlnSerTyrLeuGlnAlaGlyValIleSerTyrGly 251
97278 TGCCCTGTGGGCTGAGTCTGGCTGAGGCGGGGGTGAATCACTGGGTA 97229
251 uGlyCysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAla 268
97228 GGGCTGTCCCGCCAGAACCGCCAGGTGTCTACATCGTGTCAACGCC 97179
268 LysHisAsnTyrIleHisArgIleLeuProLysLeuGlnPheGlnProAla 284
97178 ACCCAACTGGATTCATCGATCATCCCAAACTGACATTCACACGACG 97129
285 ArgLeuGlyGlyGlnLys 290
97128 AGGTTGGCGGCCAAGAG 97111
seq_name: gb_hgt:AC092117
seq_documentation_block:
LOCUS AC092117 208529 bp DNA linear HTG 23-AUG-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-227P14, WORKING DRAFT
SEQUENCE. 2 ordered pieces.
ACCESSION AC092117
VERSION AC092117.2 GI:15281234
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208529)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
REFERENCE 2 (bases 1 to 208529)
DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (22-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 23, 2001 this sequence version replaced gi:14522963.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have

* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 154083: contig of 154083 bp in length
* 154084 154183: gap of unknown length
* 154184 208529: contig of 54346 bp in length.

FEATURES
Source 1.208529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CTD-2270P14"

BASE COUNT 48217 a 56176 c 56702 g 47334 t 100 others
ORIGIN

ment_scores:
Quality: 997.00 Length: 1090
Ratio: 3.734 Gaps: 4
Percent Identity: 24.495

alignment_block:
US-10-041-006-7 x AC092117 ..

Align seg 1/1 to: AC092117 from: 1 to: 208529

```
22 ALAAlaThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyI 38
|||||
105010 GCTCCCCAGCCTGTGGTGGCCCGGATGCTGAACCGAATGGGGGCGG 105059
38 YGLAspThrGlnGluGlyIuTrpProTrpGlnValSerIleGlnArgA 55
|||||
105060 GCAAGACACGAGAGAGGCGAGTGGCCCTGGCAAGTCAGCATCCAGCGCA 105109
55 snGlySerHisPheCysGlyGlySerLeuIleAlaGluGlnTrpValIleu 71
|||||
105110 ACGGAGGCACTTCTGGGGGGGCAAGCTCATCGGAGCAAGTGGGCTCTG 105159
72 ThrAlaAlaHisCysPhe..... 77
|||||
105160 ACGGCTGGCAGCTGCTCCGCAAGTGAATCCGCCGCCCTGCCCCGCC 105209
77 ..... 77
105210 CATAGCGCTGACAGCGCCCGCGCGAGCGGTTCAAGACCGGTGAGCA 105259
77 ..... 77
105260 GCCGCGCGCGCAATCCTGGGGGTGAACCTCCCTGGGGGCTCTGTTCCA 105309
77 ..... 77
105310 GCCCCTCCACCAAGATGCTTCCTTAGTCCAACTCCAGGGCTAAGTTC 105359
77 ..... 77
105360 CAGTTGCAACCGCTGCTCCGCCCGGAGAGTGCCTGCACCGCCCCC 105409
77 ..... 77
105410 GACCCCTCATTCCTCCCTCACCACCACTCACCTCCCTGTGGTCCCTGC 105459
77 ..... 77
105460 AGAAGCGCGCGGAGGCTGTGCCACCGCGCCCTCTGGCTTTCCCA 105509
77 ..... 77
105510 TCCCGCACACCTCAGCTCCAGGACACTTTCGGGAGGAACTGTGCT 105559
77 ..... 77
```

```
105560 CACAAAGCCCAAGAGACAGACAGAAAGGCCCTTCTCCCTCACCACCT 105609
77 ..... 77
105610 GAACCAACCCCAAGAAAGCCCTGAGAGAGCGAGCCACCCAGCCCTGTC 105659
77 ..... 77
105660 CATGTATGACCACTGTGTCACACACTTCCGGGTGTCCAGGCCCTTC 105709
77 ..... 77
105710 AACTCAGACCTCAACACGCGAGCTCAATTATTTAAACCCACATCTTT 105759
77 ..... 77
105760 TTCTTTTCTTCTTCTGATCTTTAAAGAAATGATGACAAAAAAC 105809
77 ..... 77
105810 CCACATCTTAATTCAGATACACAGGCGAGGACGGTGGCTCACACCG 105859
77 ..... 77
105860 TAATCCAGCACTTTGGAGGCCAAGCGGCGAGATCAAGTGAGCCAG 105909
77 ..... 77
105910 AGTTCAAGACAGCCCGGGCGACACAGAACACCTGTCTACTTAAAA 105959
77 ..... 77
105960 TAAAAAATTAATAAACAAAAAACAAAAAACAAAAAACCCAGTGC 106009
77 ..... 77
106010 CAGTGTGCTGCTGTGTCACCAAGTACTCAGAAAGTGTGAGAGAGTGC 106059
77 ..... 77
106060 CTTGAGCCAGAGAGTGAAGGCTGAGTGAAGTCACTGACCTGCTGCA 106109
77 ..... 77
106110 CTCAGGCTGGGTGACAGATGGAATCTGTCTTAAGGGGAAAAA 106159
77 ..... 77
106160 ATCGAAACTCGGGGTGAGAGAGGCGAGTCACTGCGAGAGCCCGCAG 106209
77 ..... 77
106210 GTGGGGCTGGCCCTGCTGACCTGAGAGGTGTTCTTCCGAGCAGCT 106259
77 ..... 77
106260 TTGAGTAGGCACTGTGACTGTCCCAACAGCAAGGTGAGAAAGTGCAT 106309
77 ..... 77
106310 CACAGAACTGTTAAGTACTTGCCCAAGGCCACACAGCGCGCAGGG 106359
77 ..... 77
106360 TGGAGTGGAATCTGAGTGGCGGCTGTGGCTTCAATGTCAACAGCAT 106409
77 ..... 77
106410 CACCGTGCACCTGGCTGTCTAGAGATGACTCTGGAAGACATTTGCA 106459
77 ..... 77
106460 AGAGAGGTGTGGGCGACAGGTGTGGCCAGAGAGCCCGACCCCTCCCG 106509
```

78Ar 78
106510 GCCTGTCACACAGAGGGCCGCTCAGCCGCTCTCTGTTTCCCG 106559
78 g.AsnThrSerGluThrSerLeuTyrGlnValLeuLeuGlyAlaArgGln 94
106560 CCAGCACTCTGAGAGCTCCCTGTACAGAGTCTGCTGGGGGAGAGCAG 106609
95 LeuValGlnProGlyProHisAlaMetTyrAlaArgValArgGlnValG 111
106610 CTAGTGAGCGGAGACACAGCTATGATGCCCCGGGTAGGAGGTGA 106659
111 userAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeu 128
106660 GAGCAACCCCTGTACAGGGGACAGGCTCCAGGCTGAGCTGGCCCTG 106709
128 a.lGlueGlnAlaProValProPheThrAsnTyrIleLeuProValCys 144
106710 TGGAGCTGGAGGACACAGTGCCTTACCAATTCATCTCCCGGTGTC 106759
145 LeuProAspProSerValIlePheGluThrGlyMetAsnCysTrpValTh 161
106760 CTGCTGACCCCTCGGTGATCTTGAAGCGGCGATGATGCTGGGTGTC 106809
161 rglYTrpGlySerProSerGluGlu..... 169
106810 TGGCTGGGGGAGCCCGCAGTGAAGAGTAAGGGACAGGGCTGGGAAGA 106859
169 169
106860 ATGGGGGATATGCTCTCAAGAGGCCAGAGCGGCTTGGGGACACTGTGA 106909
169 169
106910 ACCCACAATCTCTGTGTGCTATATTAATACATACACAGGCTCTGTTC 106959
169 169
106960 ACACAGGTGGGCCGAGGGAGCAGCAGCAGACCCAGAAAGAAAGGGG 107009
169 169
107010 TGAAGGAGAGGGGAGACAGAAAGGCCAGCCTGCTGTGAAGAGAGG 107059
169 169
107060 AGGATGTCGCCCCACAGTGCCTTAAGCCAGAGATGTCTAAACATGAGG 107109
69 169
110 AAACGACAGGAGAGATGAGTGTGTCCCGGGCCGTGGGTCTGTGAT 107159
170AspLeuLeu 172
107160 GAGAGATGCTGTTGAGCCCTGGGCTGTTCCGCCCTCTCCACAGCTCCG 107209
173 ProGluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrPr 189
107210 CCGAAGCCGCGGATCTCTGAGAACTGCTGTGCCATATGACACACAC 107259
189 olyScyAsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProL 206
107260 CAAGTGCACACGCTCTACAGCAAGACACCGAGTTGGCTACCAACCA 107309
206 ystThrIleLysAsnAspMetLeuCysAlaGlyPheGlnGlnLysLys 222
107310 AAACCTTCAAGATGACATGCTGTGTGCGCGGCTTCAAGAGGGCAAGAG 107359
223 AspAlaCysLys..... 226
107360 GATGCTGCAAGGTGGAGCAATGTGTGTCTCATGAGGAGCTCAGTCCCA 107409

226 226
107410 CCGCAGGCCCCAGGGCAGGTGGATCATGCCCGCTCCATGGGCTATGC 107459
226 226
107460 ATTCAACAGGTGACACAGCTTTGTTTGAAGGAGAGATTCCAGGGCCGG 107509
226 226
107510 TGTGTGCTCAGCGCTGTATTCAGACACTTTGGAGGCCAAGAGAGGT 107559
226 226
107560 GSATACAAAGGTCAAGNATGAGACCATCTCTGGCTTAACACAGTGAACC 107609
226 226
107610 CCATCTCTACTAAAAATACAAAAAATTAGCCAGGCGTGTGTGGCGCC 107659
226 226
107660 TGTACTCCAGCTACTTGGAGGCTGAGGCGAGAGAAATGGCGGAACTG 107709
226 226
107710 GSAGGCGAGGCTTGAGTGAAGCCAGATCGGCGCACTGCACCTCAGCTG 107759
226 226
107760 GGCCACAGAGACAGCTCCGTCTCATATMAAAAAAAAAAGAAAGCAGA 107809
226 226
107810 TTCTGAGCCGCCACCCAGAAAGATCAGAGTCAATGGCTGCGGCTGGC 107859
226 226
107860 CACAAGAACTGCATTTTCAACACACACACACACCTAGTTGATCTGGGG 107909
226 226
107910 ACACCTAAGATAGGCACTTTAGACAAATCGGACTGGAGGGGATGGGAG 107959
226 226
107960 CCGGCGAGAGGCTGTGGGATGAGAGAAAGAAATCCATTGGGAGCTAC 108009
226 226
108010 ACACCTCTGTGCGGCTGCAACTGAGTCTGTGGGAACTCAAGTGGGCTC 108059
227GlyAspSerGlyGlyProLeuVal 234
108060 TGACGTGACTCTGCTGCTCCCGGAGGGCAGCTGGGCGGCCCTGTGTG 108109
235 CysLeuValGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGly 251
108110 TGCTCTGGGTGCTAGTGTGTGCTGCTGAGCGGGGTGATACGTGGGTGA 108159
251 uGlyCysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaH 268
108160 GGGCTGTGCCCGCAGAACCGCCAGGTGTGTACATCTGTCACCGGCC 108209
268 sHisAsnTrpIleHisArgIleIleProLysLeuGlnPheGlnProAla 284
108210 ACCAACTGGATCATGAGATATCCCAACTGAGATTCCAGCAGCGC 108259
285 ArgLeuGlyGlyLys 290
108260 AGTTGGGCGGCCAGAG 108277
seq_name: gb_pr:AC004036

```

seq documentation block:
LOCUS AC004036 40179 bp DNA linear PRI 27-JAN-1998
DEFINITION Homo sapiens chromosome 16, cosmid clone 400DI (LANL), complete
sequence.
ACCESSION AC004036
VERSION AC004036.1 GI:2811104
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 40179)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maitblie,M.,
Mitra,M. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
Unpublished
2 (bases 1 to 40179)
Ricke,D.O. and Wagner,R.P.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) system
Unpublished
3 (bases 1 to 40179)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maitblie,M.,
Mitra,M. and Deaven,L.
Direct Submission
JOURNAL Submitted (27-JAN-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1.40179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="400D1"
1.92
/rpt_family="AluJo"
complement(547..899)
/rpt_family="Alu"
complement(900..1124)
/rpt_family="AluSx"
1180..1463
/rpt_family="Alu"
1479..1764
/rpt_family="Alu"
/rpt_family="Alu"
complement(1806..2099)
/rpt_family="Alu"
complement(2178..2464)
/rpt_family="Alu"
complement(2875..3215)
/rpt_family="Alu"
3996..4294
/rpt_family="Alu"
4578..4836
/rpt_family="Alu"
4898..5098
/rpt_family="Alu"
/rpt_family="Alu"
complement(5129..5463)
/rpt_family="AluSx"
complement(5449..5798)
/rpt_family="Alu"
complement(5869..6176)
/rpt_family="Alu"
6439..6779
/rpt_family="Alu"
6800..7118
/rpt_family="Alu"

```

```

/rpt_family="Alu"
complement(7280..7602)
/rpt_family="Alu"
7288..7307
/note="(r)20"
/rpt_type=tandem
/rpt_unit="r"
complement(7651..7977)
/rpt_family="Alu"
complement(7949..8271)
/rpt_family="Alu"
complement(8469..8662)
/rpt_family="AluSx"
8782..9102
/rpt_family="AluSx"
9089..9112
/note="(AAT)8"
/rpt_type=tandem
/rpt_unit="AAT"
9193..9329
/rpt_family="L1"
9260..9654
/rpt_family="Alu"
9801..10106
/rpt_family="MLT1"
10245..10303
/rpt_family="Alu"
complement(11162..11322)
/rpt_family="Alu"
11357..11733
/rpt_family="Alu"
complement(11734..12016)
/rpt_family="Alu"
complement(12160..12264)
/rpt_family="Alu"
12313..12638
/rpt_family="Alu"
complement(12840..13147)
/rpt_family="Alu"
complement(13345..13714)
/rpt_family="Alu"
14679..15069
/rpt_family="Alu"
16880..17196
/rpt_family="Alu"
17175..17194
/note="(A)20"
/rpt_type=tandem
/rpt_unit="A"
19639..19803
/note="SP0031A Serine Protease exon"
19727..19792
/note="SP0031A Serine Protease exon"
20443..20771
/rpt_family="Alu"
complement(20874..21003)
/rpt_family="MIR"
21239..21463
/note="SP0031A Serine Protease exon"
21252..21446
/note="100% identity EST124219 CAMALBe cell line II"
22114..22420
/rpt_family="Alu"
complement(22411..22493)
/rpt_family="MERS"
22692..22879
/note="SP0031A"
complement(22899..23084)
/note="98% identity EST80292"
/db_xref="dbEST:AA368960"
24791..24832
/note="(AC)21"

```


78 78
20970 GTGAGGTGAAATCTGATCGGGGCTGTGCTTCATGTCAACACGCA 21019
78 78
21020 TCACCGCTGCACCTGGCTGTCTTACAGATGACTCGAAGACATTGTGCA 21069
78 78
21070 GAGAGAGGTGTGGGGCAGGTGTGGCCAGAGAGACCCACGCCCTCCG 21119
78 78
21120 GGCCTGTCCACACAGAGGGGCCCTCAGCGCCCTCTCTGTTCCTCC 21169
79 .Asn...ThrSerGluThrSerLeuTyrGlnValLeuLeuGlyAlaArgI 94
170 GCCAGCACCCTGAGAGCGTCCCTGTACAGGTCCTGTGGGGGCAAGCA 21219
94 nIeuValGlnProGlyProHisAlaMetTyrAlaArgValArgIValG 111
21220 GCTACTGCAGCGGGGACACACAGCTATGTATGCCGGGTGAGGCGGTGG 21269
111 LuSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeu 127
21270 AGAGCAACCCCTGTACAGGGGACGCGCTCCAGCGCTGAGCGTGGCCCTG 21319
128 ValGluLeuGluAlaProValProPheThrAsnTyrIleLeuProValIcy 144
21320 GTGGAGCTGGAGGACACAGGCTTCACCAATTCATCTCCCGTGTG 21369
144 sIeuProAspProSerValIlePheGluThrGlyMetAsnCysTrpValT 161
21370 CCTGCTGACCCCTCGGTGATCTTGAACGCGCATGAACCTGTGGTCA 21419
161 hGlyTyrPglySerProSerGluGlu..... 169
21420 CTGGCTGGGGGACGCCAGTGAAGGTAAAGGGACAGGGCTGGGAAG 21469
169 169
21470 AATGGGGATATGCTCAAGAAAGGCCAGAGCGCTTGGGACCACTGTG 21519
169 169
21520 AACCCACAATCTCTCTGTGTATTAATACATACACAGGTCTCTGT 21569
169 169
21570 CACACAGTGGGGCGAGGGGAGCAGACAGACCAGAAAGAAAGAGGG 21619
169 169
21620 GTGAAGGAGAGGGGACAGAAAGGCCAGCCTGCTGCTGGAAGAGG 21669
169 169
21670 GAGGATGTCTGCCCAAGTGCCTAAGCCAGAGATGTGCTAAACATGAG 21719
169 169
21720 GAAAGCAGAGCAGATGAGTGTGTCCCGGGCGGTGGTTCTTGA 21769
170 AspLeuLe 172
21770 TAGAGAACTCCTTGAGCCCTGGCTGTCCGCCCTCTCCAGACTCTCT 21819
172 uProGluProArgIleLeuGlnIleValProIleIleAspTrp 189
21820 GCCGAAACCGCGATCTGTGCAAACTCGCTGTGCCCATTCACGACACAC 21869
189 rOlyCysAsnLeuLeuTyrSerLysAspTrpGluPheGlyTyrGlnPro 205

|||||
21870 CCAAGTCAACCTGCTCTACAGCAAAAGACAGGATGTGGCTACCAACC 21919
206 LysThrIleLysAsnAspMetLeuGlyPheGluLeuLysLys 222
21920 AAAACCATCAAGAAATGACATGTGTGCGCGGCTTCAGAGAGGCAAGAA 21969
222 sAspAlaCysLys..... 226
21970 GGATGCTCGCAAGGTGGAGCATGTGTGTCCATGGGAGCTAGTCCC 22019
226 226
22020 ACCGCAAGCCCAAGGCAAGGTGGATCATGCCCGCTCCATGGGCTATG 22069
226 226
22070 CATTCAAACAGCTGAGCAGCTGTTTGAAGAAAGCAGATTCCAGGGCCGG 22119
226 226
22120 GTGTGTGGCTCACGCGCTGTATCCAGCACTTTGGGAGGCCAAGAGGG 22169
226 226
22170 TGGATCACAAGGTCAGAGATGAGACCATCTGGCTAACACAGTGAAC 22219
226 226
22220 CCCATCTCTACTAAATAATACAAAAAATTAGCAGCGGTGTGGGCCG 22269
226 226
22270 CTGTAGTCCACGACTACTTGGGAGCTGAGGACAGAGAATGGCCGAAC 22319
226 226
22320 GGGAGCGGAGCTTGCAGTAGCGAGAGTCGCGCAGTGCACCTCAAGCT 22369
226 226
22370 GGGCCACAGACAAAGACTCCGCTCAATAAAAAAAGAAAGAACAG 22419
226 226
22420 ATTCTGGACCCGCCACAGAGATCAGATGAGTGTGTGGGTGGG 22469
226 226
22470 GCACAGAAACTGCATTTTCAACACACACACACACCTAGTGTATGAGG 22519
226 226
22520 GACACTAAGAAATAGGCACTTTAGACAAATCGACTGCGAGGGATGGGA 22569
226 226
22570 GCCGGGAGAGAGGCTGTGGATGAGAGAGAAGAAATCATTTGGGAGCTA 22619
226 226
22620 CACACCTCTGTGGCGCTGCACACTGAGTCTCTGGGAACCTCAAGTGGGCT 22669
227 GlyAspSerGlyLysProLeuVal 234
22670 CTGAGGCTGACTGCTGCCGCCCGCAGGCGGACTCGGGCGGCCCTGGT 22719
234 LysLeuValGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGlyC 251
22720 GTGCTGTGGGTCACTGTGTGCGACGCGGGGGGTGATCAGCTGGGGTG 22769
251 LuGlyCysAlaArgGlnAsnArgProGlyValIleArgValThrAla 267
|||||

55 snGlySerHisPheCySGlyGlySerLeuIleAlaIleGlnTrpValLeu 71
 |||||
 5038 ACGGAGCCACTCTGCGGGGAGCCCTCATCGCGGAGCATGGGCTCG 5087
 72 ThrAlaIleHisCySphe..... 77
 |||||
 5088 ACGGCGCGCACTGCTCCCAAGTAGTCCGCCGCCCTGCCCCCGCC 5137
 77 77
 5138 CATAGCGTGAAGCGCCCCCGCGCTACCGGTTACAGCACCGTGACAGC 5187
 77 77
 5188 GCCCGCGCGCCAAATCTGCGGGTAGCTTCTGTGGGGCTCTGTGCA 5237
 77 77
 5238 GCCCTCCCAACCATGCTTCCCTTAGGTCCAACTCAGGGCTAACTTC 5287
 77 77
 5288 CAGTTGCACACCGGTGTGCCCGCGGGAGGTGCTCGCACCGTCCGCC 5337
 77 77
 5338 GACCCCTTCATCTCTCCACCCACTACCCACTCTGTGGGTCCCTGC 5387
 77 77
 5388 AAGACGGCGCGCGGCTCTGCCACCGGCCCTGTGGCTTTCCCA 5437
 77 77
 5438 TCCCGCACACACTCAGTCCAGACACTCTCCGGAGAGACTCTGCT 5487
 77 77
 5488 TACAAAGCCCAAGACAGACAGAACGGCCCTCTCCCTCACCCACT 5537
 77 77
 5538 GAAACACCCCAAGAACCCCTGAGCAGAGGCCAGCCACCCCTCTGC 5587
 77 77
 5588 CATGTATGAACACCTGGCCCCACACTTCCGGGTGTCCAGGCCCTCA 5637
 77 77
 5638 CCTCACACCTACACCGCAGCTTAATTATTAAACCCACACTCTTT 5687
 77 77
 5688 TCTTTTTTCTCTTGATCTTTAAAGATATCATGACAAAAAACCC 5737
 77 77
 5738 CACATCTTAATTCAGATACTACGGCCAGCAGGTGGTCTACACCGCT 5787
 77 77
 5788 AATCCACACACTTTGGAGGCCAAGCGCGGAGATCAGTTGAGCCAGGA 5837
 77 77
 5838 GTTCAGACAGCCCGGCAACAGACAGACCTGTCTACTAAAT 5887
 77 77
 5888 AAAAAAAAAATTAAAAACAAAAACAAAAAACCCAGTGAGTG 5937
 77 77

5938 GTGCTGTGGTCCCAAGTACTCAGGAAGCTGAGCGAGAGATGGCTTGA 5987
 77 77
 5988 GCCAGAGATTGAGAGGCTGAGTAGAGTACAGTTGCAACACTGCATCCA 6037
 77 77
 6038 GCTGGGTGACAGATGAAGAACTGTCTTAGGGGAAAAAATGGA 6087
 77 77
 6088 AACTGGGCTAGAGAGAGGAGTCACTGCCAGAGCCGAGTGGG 6137
 77 77
 6138 CTGGCCCTGTGCACCTGAGAGGTGTCTCTCCGAGCAAGCTTTCAGG 6187
 77 77
 6188 TAGCACTGTACTGTCCCAACAGCACAGGTGAGAAAAGTGCATCAGA 6237
 77 77
 6238 GAAGTTGTTAAGTACTTGCCCAAGGCCACAGCGGCGAGGGTGGAGG 6287
 77 77
 6288 TGAATTCGAGTGGGCGGTGTGCTTCAATGCACACTCATACCGC 6337
 77 77
 6338 TGCACCTGGCTCTCTAGAGTGGCTCTGAGACACTGTGAGAGAG 6387
 77 77
 6388 GTGTGGGCAACAGTGTGGCAGAGAGTCCCGCCCTCCGGGCTGT 6437
 78 Arg, AsnT 80
 6438 CCCCACAGAGGGCCCTACAGCCGCTCTCTGTTTCTCCGCGACGA 6487
 80 hrSerGluThrSerLeuTyrGlnValLeuLeuGluAlaArgGlnVal 96
 6488 CTTGTGAGAGTCCCTTACAGGTCCTGTGGGGCAAGGCAAGTGTG 6537
 97 GluProGlyProHisAlaMetTyrAlaArgValArgGlnValGluSerAs 113
 6538 CAGCGGGACACACGCTATGTATGCCGGGTGAGGGGTGAGAGCAA 6587
 113 nProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaLeuValGlu 130
 6588 CCCCCTGTACAGGGCATGGCTTCAGCCGTGACGTGGCTGTGTGAGC 6637
 130 euGluAlaProValProPheThrAsnTyrIleLeuProValCysLeuPro 146
 6638 TGGAGGACCAAGTCCCTTACCAATTACTCTCCCGGTGCTGCT 6687
 147 AspProSerValIlePheGluThrGlyMetAsnCysTyrValThrGlyTr 163
 6688 GACCCCTCGGTGATCTTTGAGAGCGGCATGAACTGCTGGGTCACTGGC 6737
 163 pGlySerProSerGlu, Glu..... 169
 6738 GGGAGCCCAAGTAGAGAGTAAGGAGCAGGGCTGGGAAGAAATGGG 6787
 169 169
 6788 GATGTGCTCAAGAGGCCAGGAGGCTTGGGAGCACTGTGAACCCAC 6837
 169 169

```

6838 AATCTCTGCTGTAATTAATACATACACAGGCTCTGTTACACAG 6887
169 ..... 169
6888 GTGGGGCCGAGGGGAGCAGCAGACCCAGAAAGAGGGGTGAAG 6937
169 ..... 169
6938 GAGAGGGAGAGAAAGGGCCAGCTGCTGCGAAGAGGAGATGT 6987
169 ..... 169
6988 CTGCCCCACAGTGCCTTAAGCCAGATGCTTAACATGAGAGAGC 7037
169 ..... 169
7038 AGCAGAGTGGAGTGTGTCCCGGGGCGGTCTTGATGAGAGAG 7087
170 ..... AspleuLeuProgluP 175
7088 TCCGTTAGCCCTGGGCTGTTCGCCCTCTCCAGACTCTCGCCGAA 7137
175 roArgileuGlnLysLeuAlaValProIleAspThrProLysCys 191
|||||
7138 CGCGATCTTCAGAAATCGCTGTGCCATCATCAGACACCAAGTGC 7187
192 AsnLeuLeuTyrSerLysAspThrGluPheGlyTyrGlnProLysThr 208
|||||
7188 AACCTGCTCTACACAAAGACACGAGTTGGCTACCAACCAACAT 7237
208 eLysAsnAspMetLeuCysAlaGlyPheGluGlyLysLysAspAla 225
|||||
7238 CAAGAAATGACATGCTGTGCCGCTTCGAGAGGCAAGAGATGCT 7287
225 ySLys..... 226
7288 GCAAGGTGGAGCGATGTGTCCATGGGACTCCTCACCGCAG 7337
226 ..... 226
7338 GCCCAGGGCAGGTGGATCATGCCCGCTCATGCGCTATGCTTAC 7387
226 ..... 226
7388 CAGTGAACACCTTGTGTAAGAGGAGATTCCAGGGCGGCGGCTG 7437
226 ..... 226
7438 GCTCAGCCCTGTAAATCCAGACTTGGGAGGCAAGAGGGTGAATC 7487
226 ..... 226
7488 AAGGTGAGAGATGAGACCATCTGGCTAACACAGTGAACCCGCTC 7537
226 ..... 226
7538 TACTTAAATATCAAAAAATTAAGCAGCGGTGTGTGGCGCTGTATC 7587
226 ..... 226
7588 CCAGCTACTCGGAGGCTGAGGAGAGAGATGGCTGAACCTGGAGGCG 7637
226 ..... 226
7638 GAGCTTGACAGTGAAGATCGCCGCTGACACTCCAGCTGGGCCACA 7687
226 ..... 226
7688 GAGCAAGATTCGCTCAAAAAAAAAAGAAAAAGAACAGATTCC 7737
226 ..... 226
7738 TGGACCCACCCAGAGAGATCAGTCAGTGGGTCTGGGGTGGGCACA 7787

```

```

226 ..... 226
7788 AGAAATGCAATTTTCACACACACACACACCTTATGATCTGGGAGAC 7837
226 ..... 226
7838 TAAAGATATGACATTAGACAATCGACTGGCAGGGGAGCGGAGCCG 7887
226 ..... 226
7888 GCAGAGAGCTCTGGAGAGAGAGAAATTCATTGTGAGCTACACGC 7937
226 ..... 226
7938 CTCTGTGCGCTGCACTGAGTCTCTGGAACTCAAGTGGGCTTAC 7987
227 ..... GlyAspSerGlyLysProLeuValCysL 236
7988 GGTGATTTGCGTCCCGCAGGCGGAGTGGGCGGCGCTGTGTGCC 8037
236 euValGlyLysThrLeuGlnAlaGlyValIleSerThrGlyLys 252
|||||
8038 TCGTGTGCTGAGTGTGCTGCGAGCGGGGTATCAGCTGGGTGAGGC 8087
253 CysAlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHis 269
|||||
8088 TGTGCGCGCAGAACCGCCAGTGTATACATCGCTGACCGCCACCA 8137
269 sAsnThrIleHisArgIleIleProLysLeuGlnPheGlnProAla 286
|||||
8138 CAACGTGATCATCGATCATCCCAAACTGCAGTTCAGCCAGCAAGT 8187
286 euGlyGlyGlnLys 290
8188 CGGGCAGCCAGAG 8201
seq_name: gb_pat:AX001347
seq_documentation_block:
LOCUS AX001347 683 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 3 from Patent EP0890646.
ACCESSION AX001347
VERSION AX001347.1 GI:7241523
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 683)
AUTHORS
Burgess,N.A. and Southan,C.D.
TITLE
HE2NM40 Serine Protease
Patent: EP 0890646-A 3 13-JAN-1999;
JOURNAL
SMITHKLINE BEECHAM PLC (GB)
FEATURES
location/Qualifiers
1..683
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 156 a 205 c 195 g 119 t 8 others
ORIGIN
alignment_scores:
quality: 810.50 length: 220
ratio: 4.243 gaps: 7
Percent Similarity: 86.818 Percent Identity: 79.091
alignment_block:
US-10-041-006-7 x AX001347
Align seg 1/1 to: AX001347 from: 1 to: 683
78 ArgAsnThrSerGlnThrSerLeu.TyrglnValLeuGlnAlaArg 94
|||||

```



```

1 CGCAACACCTGTGAGACGTCCCTGGACACAGCTCTGCTGGGGCGAAGGC 50
94 InLeuValGlnPro.GlyProHisAlaMetTyrAlaArgValArgGlnVal 110
51 AGCTAGTGCAGCCGGGGACACACACCTGTATGTATGCCGGGTGAGGCAAGT 100
110 IGlusSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127
101 GGAGAGCAACCCCTGTACAGAGGCGCTCCAGCGCTGACGTGGCC 150
127 euValGlnLeuGlnAlaProValProPheThrAsnTyrIleLeuProVal 143
151 TGGTGGAGCTGGAGGACACAGTCCCTTACCAATTCATCTCTCCCGCTG 200
144 CysLeuProAspProSerValIlePheGlnThrGlyMetAsnCysTyrVal 160
201 TGCCTGCTGACCCCTCGGTGATCTTTGAGACGGGCAATGAACTGTGGGT 250
160 IThrGlyTyrGlySerProSerGlnGluAspLeuAspProGluProArgI 177
251 CACTGGCTGGGGGACCCCGAGTGAAGAAAGACCTCTGCGCGGACGGGGA 300
177 IeLeuGlnIlyLeuAlaValProIleIleAspThrProLysCysAsnLeu 193
301 TCCCTCAGAACTCCGTGTGCCATCATCGACACACCCAGTGCACCTG 350
194 LeuTyrSerLysAspThrGlnPheGlyTyrGlnProLysThrIleLysAs 210
351 CTCTACAGCAAGACACGAGTTTGGCTACCAACCCAAACCATTAAGAA 400
210 nAspMetLeuCysAlaGly.PheGlnGlnGlyLysLysAspAlaCysLys 226
401 TGACATGCTGTGCGCGGNTTTCAGAGAGGCGCAAGAGATGCTCGAAG 450
227 .GlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerTyrLeu 243
451 TGGG.....GGAGCCATGTGGTGTCCATGGGATTCAGTTCACAC 491
243 InAlaGlyVal...IleSerTyrGlyGlnGly.....Cys 253
492 CGNCAGGCCCGAGGCGAAGGTGGGTAAAGCCCTGTCCATGGGTATG 541
254 AlaArgGlnAsnArgProGlyValTyrIleArgValIleThrAlaHisLys 270
542 CTTCAAAGCAAGNCCGCCAGTGTAAATCCGTGTAAGCGCCACAA 591
270 nTrrPheHisArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287
592 CTGGATCCATCGATCATNCCCAACTGAG..TTCAGGCCAACGAAGTTTG 640
287 Iy 287
641 GC 642

```

seq_name: gb_pat:E31783

seq_documentation_block:

LOCUS E31783 683 bp DNA linear PAT 07-FEB-2001

DEFINITION Novel compound.

ACCESSION E31783.1 GI:13026587

VERSION JP 19908992-A/2.

KEYWORDS unidentified.

SOURCE unidentified.

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 683)
NICOLO, A.B.H.H. and KurlinkembIado, C.D.S.S.ADTHONS Novel compound
Patent: JP 19908992-A 2 13-APR-1999;

JOURNAL SMTHTLINE BECHAM CORP PUBLIC LTD CO

COMMENT OS Unidentified
PN JP 19908992-A/2
PD 13-APR-1999

```

PF 09-JUN-1998 JP 1998199425
PR 10-JUN-1997 GB 9712088.5,17-OCT-1997 GB 97308295-1, PR
20-FEB-1998 GB 9803650.2
PI NICOLO ANN BAJESU HELEN ELIZABETH KURLINKENBIADO, PI
CHRISTOPHER DONALD SAZAN
PC C12N15/09,A61K31/70,A61K35/76,A61K38/00,A61K38/00,
PC A61K39/395,
PC A61K39/395,A61K48/00,C07K16/40,C12N5/10,C12P21/02,C12P21/08,
PC C1201/68,
PC C12N15/00,A61K37/02,A61K37/02,A61K37/02,C12N5/00 CC
Strandedness: Single;
CC Topology: linear;
FH Key 1 Location/Qualifiers
FT source 1.683 /organism='Unidentified'.
FEATURES
source Location/Qualifiers
BASE COUNT 156 a 205 c 195 g 119 t 8 others
ORIGIN

```

```

alignment_scores:
  Quality: 810.50 Length: 220
  Ratio: 4.243 Gaps: 7
Percent Similarity: 86.818 Percent Identity: 79.091

```

alignment_block:

US-10-041-006-7 x E31783 ..

Align seg 1/1 to: E31783 from: 1 to: 683

```

78 ArgAsnThrSerGluThrSerLeu.TyrGlnValIleLeuGlnValArgG 94
|||||
1 CGCAACACCTGTGAGACGTCCCTGGACACAGTCTGCTGGGGCGAAGGC 50
94 InLeuValGlnPro.GlyProHisAlaMetTyrAlaArgValArgGlnVal 110
51 AGCTAGTGCAGCCGGGGACACACACCTGTATGTATGCCGGGTGAGGCAAGT 100
110 IGlusSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127
101 GGAGAGCAACCCCTGTACAGAGGCGCGCTCCAGCGCTGAGGTGGCC 150
127 euValGlnLeuGlnAlaProValProPheThrAsnTyrIleLeuProVal 143
151 TGGTGGAGCTGGAGGACACAGTCCCTTACCAATTCATCTCTCCCGCTG 200
144 CysLeuProAspProSerValIlePheGlnThrGlyMetAsnCysTyrVal 160
201 TGCCTGCTGACCCCTCGGTGATCTTTGAGACGGGCAATGAACTGTGGGT 250
160 IThrGlyTyrGlySerProSerGlnGluAspLeuAspProGluProArgI 177
251 CACTGGCTGGGGGACCCCGAGTGAAGAAAGACCTCTGCGCGGACGGGGA 300
177 IeLeuGlnIlyLeuAlaValProIleIleAspThrProLysCysAsnLeu 193
301 TCCCTCAGAACTCCGTGTGCCATCATCGACACACCCAGTGCACCTG 350
194 LeuTyrSerLysAspThrGlnPheGlyTyrGlnProLysThrIleLysAs 210
351 CTCTACAGCAAGACACGAGTTTGGCTACCAACCCAAACCATTAAGAA 400
210 nAspMetLeuCysAlaGly.PheGlnGlnGlyLysLysAspAlaCysLys 226
401 TGACATGCTGTGCGCGGNTTTCAGAGAGGCGCAAGAGATGCTCGAAG 450
227 .GlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerTyrLeu 243
451 TGGG.....GGAGCCATGTGGTGTCCATGGGATTCAGTTCACAC 491

```

```

243 InAlaGlyVal...IleSerTrpGlyGlyGly.....Cys 253
492 GCGNCGCCGACGAGGCGGTGAGGCCCTGTCATGGGTATG 541
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHisAs 270
542 CTTCAAGAGGAGAACGCCAGGTCTTAATCCGTGTAACGGCCCAACAAA 591
270 nTrpIleHisArgIleIleProLeuValGlnPheGlnProAlaArgLeu 287
592 CTGGATTCATCGATCATNCCCAACTGAG.TTCCAGCCAAAGAAATTGG 640
287 1y 287
641 GC 642

seq_name: gb_pr:AB010779
seq_documentation_block:
  title AB010779 1306 bp mRNA linear PRI 17-MAY-2001
  description Homo sapiens hbssp-4 mRNA for brain-specific serine protease-4, complete cds.
  accession AB010779
  version AB010779.1 GI:12248750
  keywords Homo sapiens brain cDNA to mRNA.
  source Homo sapiens
  organism Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  reference 1 (sites) Mitsui,S., Okui,A., Komitani,K. and Yamaguchi,N. Cloning and characterization of a human brain-specific serine protease, hbssp-4
  journal Unpublished
  authors Yamaguchi,N. and Mitsui,S.
  title Direct Submission
  journal Submitted (27-JAN-1998) Nozomi Yamaguchi, Institute for Geriatrics, Kyoto Prefectural University of Medicine, Department of Cell Biology, Kajiki-cho 465, Kawaramachi Hirokoji, Kyoto, Kyoto 602-8566, Japan (E-mail:nozomi@koto.kpu-m.ac.jp, Tel:81-75-251-5848(ex.5848), Fax:81-75-251-5848)
  features
    source
      1. 1306
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /tissue_type="brain"
        4. .957
        /gene="hbssp-4"
        4. .957
        /gene="hbssp-4"
        /note="prosemin"
        /codon_start=1
        /product="brain-specific serine protease-4"
        /protein_id="BAB20263.1"
        /db_xref="GI:12248751"
        /translation="MVFSGAPPALGGGCLGFTSLILLASTAIIINAARIPYPACGKP OOLNRVGGEDSTDESEWPVIVSIOKNGTHCAGSLTSSVTTAAHCFEDNKNRYLF SVLGAOMOLNGSRSOKYGVAMVEHPHYSMKESACADIAIVLERSIOSEERLPI CLPDASITILPPTHCHISWGSISQDSVPLPHQOTLOKLVPTIIDEVCSHLYWRAGQ GPTEEDMLCAGTIEGRDCLDSDSGPLKCYDGAWLAGIISWEGGAERNRPVYI SLASHSWEKIVQVQYLRGAOGGALRALRAPSQSGGAARS"
    cds
      gene
      cds
  alignment_scores:
    quality: 739.00 length: 285
    ratio: 3.421 gaps: 4
    percent similarity: 75.789 percent identity: 48.772
  alignment_block:
    US-10-041-006-7 x AB010779

```

```

Align seg 1/1 to: AB010779 from: 1 to: 1306
10 LeuLeuLeuLeuCys.....PheGlySerGlnArgAlaLysAl 22
64 CTGCTGCTGCTGGCGACGACGACCATCTCAATAGCGCGCATACCTGCT 113
22 aaAlaThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyGly 39
114 TCCCCCAAGCCTGTGGAGAGCCCGACGACGTAACCGGGTTGGCGGCG 163
39 InAspThrGlnGlyGlyGlyIleuTrpProTrpGlnValSerIleGlnArg 55
164 AGGACAGCACTGACAGCAGAGTGGCCCTGATGCTGATCAGATCCAGAG 213
56 GlySerHisPheCysGlyGlySerLeuIleAlaGlnGlnTrpValLeu 72
214 GGGACCCACACCTGCGAGAGGTCTGCTCACCAGCCGCTGGGTGATCAG 263
72 fAlaAlaHisCysPheArg...AsnThrSerGluTrpSerLeuTrpGly 88
264 TGCTGCCCATGTTTCAAGACACACCTGAACAACCATACCTGTTCTCTG 313
88 alLeuLeuGlyAlaArgGlnLeuValGlnProGlyProHisAlaMetTrp 104
314 TGCTGCTGGGGCGCTGGCAGCTGGGAGACCTGGCTCTGCTGCCAGAG 363
105 AlaArgValArgGlnValGlnSerAsnProLeuTrpGln...GlyThrAl 120
364 GTGGGTGTTGCTGGGTGGAGGCCACCCCTGTCATCTCGAAGACGAG 413
120 aserSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPhe 137
414 TGCTGTGCGACATGTGCTGCTGGTGTCTGACAGCCCTCCATACGTT 463
137 hAsnTrpIleLeuProValCysLeuProAspProSerValIlePheGln 153
464 CAGAGGGGGTCTGCGCCATCTGCTACCTGATGCTATCCATCACCTCC 513
154 ThrGlnMetAsnCysTrpValThrGlyTrpGlySerProSerGlnGln 170
514 CCAACACCCACTGCTGATCTGAGCTGGGGAGGAGATCCAAAGATGAG 563
170 pLeuLeuProGluProArgIleLeuGlnLysLeuAlaValProIleLeu 187
564 TCCTTGCCGCCACCTCAGACCTGACAGAGGTGAAGTCTCATATCAG 613
187 spThrProLysCysAsnLeuLeuTrpSerLysAspThrGluPheGlyTr 203
614 ACTCGGAAGCTGTCAGCACATCTGTACTGCGGGGAGACGAGACAGGA 660
204 GlnProLysThrIleLysAsnMetLeuCysAlaGlyPheGlnGlnGly 220
661 .....CCCATTCACCTGAGGACATCTGTGTGCGGCTTGTGGAGGG 701
220 yLysLysAspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeu 237
702 GGAGCGGAGATGCTGTGCTGGGAGACGTCGGGGGCCCCCTCATGTCAG 751
237 alGlyLeuSerTrpLeuGlnAlaGlyValIleSerTrpGlyGlnGlyCys 253
752 TGGACGGCCCTGCTGCTGCGCGGATCATCAGCTGGGGCAGGGCTGT 801
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHisAs 270
802 GCGGAGCGACAGGCGCGGGCTTACATACAGCTCTTGGCCACCGGCTC 851
270 nTrpIleHisArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287
852 CTGGGTGAGAGATGCTGCAGAGGGGTGACGCTCGCGGGCGGCTCAGG 901
287 1yGly 288
|||||

```

902 GGGGT 906

seq_name: gb_pr:AF321182

seq_documentation_block:

LOCUS AF321182 1332 bp mRNA linear PRI 26-DEC-2001

DEFINITION Homo sapiens serine protease PRSS22 mRNA, complete cds.

ACCESSION AF321182

VERSION AF321182.1 GI:11386012

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

FEATURES

source

CDS

BASE COUNT

ALIGNMENT

Percent

Similarity

Ratio

Percent

Identity

Length

Gaps

Percent

Identity

Length

Gaps

Percent

Identity

Length

Gaps

Percent

Identity

Length

Gaps

Percent

228 GGGACCCACACTGGCGAGTCTCTGCTACACGCCGCTGGGTGATCAC 277

72 TALAALAHISCSphearg...AsnThrSergLutThrSerLeuTyrglnv 88

278 TGGTCCACACTGTTTCAAGACACCTGACACCAACCAACCACTGTTCTCTG 327

88 allleuleuglyAlaargGlnleuValGlnProGlyProHisAlametyrr 104

328 TGGTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 377

105 AlaargValargGlnValGlnSerAsnProleuTyrgln...GlyThrAl 120

378 GTGGGTGTGGCTGGTGGTGGGAGGCCACCTGTTATCTCGGAAGAGG 427

120 aSerSerAlaaspValAlaValGlnleuGlnAlaProValProphet 137

428 TGGCTGTCAGACATGCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 477

137 hrAsnTyrlleuProValCysleuProAspProSerValIlePheGlu 153

478 CAGAGCGGCTGCTCCATCTGCTACCTGATGCTATCCACCTCCCT 527

154 ThrGlyMetAsnCysTyrrValThrGlyTyrrGlySerProserGluAs 170

528 CCAACACACCACTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 577

170 PheLeuProGlnProArgIleleuGlnValGlnAlaValProIleleu 187

578 TCCCTGCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCT 627

187 spThrProLysCysAsnLeuLeuTyrrSerLysAspThrGlnPheGlyTyr 203

628 ACTGGAAGTCTGACCATCTGATCTGATCTGATCTGATCTGATCTGATCTG 674

204 GlnProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGluGlu 220

675CCATCACTGAGACATCTCTGTCGCGGCTGATCTGAGAGG 715

220 YLysLysAspAlaCysLysGlyAspSerGlyLysProLeuValCysLeu 237

716 GAGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765

237 aGlyGlnSerThrPheGlnAlaGlyValIleSerThrPheGlyGlyGly 253

766 TGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 815

254 AlaargGlnAsnArgProGlyValTyrrIleArgValThrAlaHisAs 270

816 GCCGAGCGCAACAGGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 865

270 nTrrIleHisArgIleleuProLysLeuGlnPheGlnProAlaArgLeu 287

866 CTGGGTGAGAGAGATGCTGCAAGGGGTGAGGCTGCGCGGCGGCTGAGG 915

287 Tygly 288

916 GGGGT 920

seq_name: gb_pr:BC009726

seq_documentation_block:

LOCUS BC009726 1403 bp mRNA linear PRI 22-OCT-2001

DEFINITION Homo sapiens, protease, serine, 22, clone MGC:9559 IMAGE:3899480,

ACCESSION BC009726

VERSION BC009726.1 GI:16307274

KEYWORDS

SOURCE

ORGANISM

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 1403)

Wed Aug 14 08:39:13 2002

us-10-041-006-7.p2n.rge

Page 21


```

34  rgmetvalgllyglinsprhngllyglutprprotrpglval 50
35  |||||||
36  GAATGGTGGCGGCGAGACGACAGAGGGAGTGGCCCTGGCAATTC 185
37  |||||||
38  SerTieglararnglyserhisphcysgllygliserleualagl 67
39  |||||||
40  AGCATTCACGCGACGAGAACCATCTTGGCGGGGACCTCATCGCGGA 235
41  |||||||
42  uglntprvalleuthralaahiscyspheararasnhrsergluthrs 84
43  |||||||
44  GCAGTGGGCTCGACGGCTGCGCACTGCTCCGCAACACTCTGAGAGT 285
45  |||||||
46  ertleutrglvalleuvalaarglnleuvalglproglpro 100
47  |||||||
48  CCTGTACAGGCTCCTGGGGGCAAGGCACTAGTGGAGCGGGACCA 335
49  |||||||
50  HisAlametyralarvalarglnvalgluserasnproleutrgl 117
51  |||||||
52  CACGCTATGTATGCCGGGTGAGGAGGTGAGAGCAACCCCTGTACCA 385
53  |||||||
54  ngllythralserserAlaspyalAlaleuvalglleuvalaapro 134
55  |||||||
56  GGGCAGCGGCTCCAGGCTGAGCTGGCCCTGGTGGAGCGAGCAG 435
57  |||||||
58  a1prophethrasntrileuaprovalcysleuProaspProserVal 150
59  |||||||
60  TGCCCTTCCCAATTCATCTCCCTGCTGCTGCTGACCCCTCGGG 485
61  |||||||
62  ATCTTGGAGAGGGGATGAATGCTGCTGCTGCTGCTGCTGCTGCT 535
63  |||||||
64  rglgluaspleuProglprobrglleuvalaavalp 184
65  |||||||
66  TGAGGAGAGCTCTGCGCAACCGGAGTCTGCGAAGTCTGCTGCT 585
67  |||||||
68  rollelleasprhprolyscysasnleuvalserlyasprhrglu 200
69  |||||||
70  CCATTCATCGACACACCAAGTGAACCTCTGACGCAAGACAGCAG 635
71  |||||||
72  PhegllytrglinProlysthrilelysasnaspmetleucysalagly 217
73  |||||||
74  TTTGGCTACCAACCCAAACATCAAGAAATGATCTGCTGCTGCTG 685
75  |||||||
76  eglugllyglinsysaspAlaCyslysglyaspsergllyglproleu 234
77  |||||||
78  CGAGGAGGCGCAGAGAGATGCTGCAAGGCGAGTGGCGGGCCCTG 735
79  |||||||
80  alcysleuvalgllyglinsertprleuglnalaglvalleasertrpgly 250
81  |||||||
82  TGGGCTCGGCGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 785
83  |||||||
84  gluglycysAlaarglnasnargprogllyvaltyrileargvalthral 267
85  |||||||
86  GAGGCTGTGCGCGCGCAAGCCCGAGGTGTACATCCGTCTCCGCG 835
87  |||||||
88  ahshsasntprilehisargtleileprolystleuglnphaglnpro 284
89  |||||||
90  CCACCAACATGATCATCATCATCATCATCATCATCATCATCATCAT 885
91  |||||||
92  laargleugllyglinsys 290
93  |||||||
94  CGAGGTGGCGCGCAGAG 905
95  |||||||

```

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS21354

seq_documentation: Drock:

ID AAS21354 standard; cDNA: 1129 BP.

AC AAS21354;

DT 24-OCT-2001 (first entry)

```

XX DE Human cDNA sequence encoding for PRO4327 polypeptide.
XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
XX KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
XX KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
XX KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO200140466-A2.
XX PD 07-JUN-2001.
XX PF 01-DEC-2000; 2000WO-US32678.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 02-DEC-1999; 99WO-US28551.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 09-DEC-1999; 99US-0170262.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 30-DEC-1999; 99WO-US31243.
XX PR 06-JAN-2000; 2000WO-US00277.
XX PR 06-JAN-2000; 2000WO-US00376.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 24-FEB-2000; 2000WO-US04914.
XX PR 24-FEB-2000; 2000WO-US05004.
XX PR 01-MAR-2000; 2000WO-US05601.
XX PR 20-MAR-2000; 2000WO-US07377.
XX PR 21-MAR-2000; 2000WO-US07532.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 17-MAY-2000; 2000WO-US13705.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 30-MAY-2000; 2000WO-US14941.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 10-NOV-2000; 2000WO-US30873.
XX (GETH ) GENENTECH INC.
XX Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
XX Geritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
XX Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX WPI; 2001-408281/43.
XX P-PSDB; AAU12282.
XX Isolated, secretory and transmembrane PRO polypeptide used to detect
XX other PRO polypeptides, link bioactive molecules to cells expressing
XX PRO polypeptides, and detect the presence of mammalian tumours e.g.
XX lung, breast, prostate, cervical.
XX Claim 3; Fig 221; 813pp; English.
XX AAS21244-AAS21518 encode for novel human secretory and transmembrane
XX PRO polypeptides. The PRO polypeptides are useful to detect other
XX PRO polypeptides, to link bioactive molecules to cells expressing
XX PRO polypeptides, to modulate biological activities of cells expressing
XX PRO polypeptides, and to detect the presence of mammalian lung, colon,
XX breast, prostate, rectal, cervical or liver tumours by comparing PRO
XX polypeptide expression in a cell sample to that in a control sample.
XX Some of the 275 sequences are also useful to stimulate the release of
XX tumour necrosis factor-alpha (TNF-alpha) from human blood, the
XX proliferation or differentiation of chondrocytes, the proliferation or
XX gene expression in pericyte cells, the release of proteoglycans from
XX cartilage, the proliferation of inner ear utricular supporting cells or
XX of T-lymphocytes, the release of a cytokine from peripheral blood

```


| | | |
|----|--------------|-----------------|
| XX | OS | Homo sapiens. |
| XX | FN | MO200153312-A1. |
| XX | XX | |
| XX | PD | 26-JUL-2001. |
| XX | XX | |
| PF | 26-DEC-2000; | 2000MO-US34263 |
| XX | XX | |
| PR | 21-JAN-2000; | 2000US-0488725 |
| PR | 25-MAR-2000; | 2000US-0552317 |
| PR | 09-JUL-2000; | 2000US-0558042. |
| PR | 19-JUL-2000; | 2000US-0620312. |
| PR | 03-AUG-2000; | 2000US-0623450. |
| PR | 14-SEP-2000; | 2000US-0662191. |
| PR | 19-OCT-2000; | 2000US-0693036. |
| PR | 29-NOV-2000; | 2000US-0727344. |

(HYSE-) HYSEQ INC.

WPI, 2001-442253/47.
P-PSDB; ANM41174.
Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
Wang J, Wang Z, Mehrman T, Xu C, Xue AJ, Yang Y, Zhang J,
Zhao Q, Zhou P, Goodrich R, Drmanac RT;

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AM3864-AM4221) with neurotrophic,
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed
specification.

Sequence 1157 BP; 220 A; 392 C; 341 G; 204 T; 0 other;

alignment_scores:

| | | | |
|---------------------|---------|-------------------|--------|
| Quality: | 1522.00 | Length: | 292 |
| Ratio: | 5.285 | Caps: | 2 |
| Percent Similarity: | 98.630 | Percent Identity: | 98.288 |

alignment_block:

US-10-041-006-7 x AAI60330

Align seg 1/1 to: AAI60330 from: 1 to: 1157

1 MetAaGpRProAlaAlaAlaProLeuLeuLeuLeuLeuCySPhaGlyse 17
90 ATAGAGCGCGCCGCGCGGTGGCTCTGCGGCGCGTGTTTTGGTC 139
27 rGlnAaGAlaLysAlaAlaAlaAlaCysGlyAaRProAaGmetLeuSna 34
140 TCAGAGCGCCAGAGCAGCAACCTGTGTGTGCCCCAGAGATGCTGAAC 189
34 rGmetAlaAlaLysAlaLysPheGlnGlnGlnGlyLutPProTArgLnaVal 50
190 GAATGGTGGCCGCGCAGACACGACAGAGGCGAGTGGCCCTGGCAAGTC 239

| | | |
|--------------------------|--|-----|
| 51 | serIleGlnArgAsnGlySerHisphesylglyserLeuIleagl | 67 |
| 240 | AGCATTCACCGCAACGGAAGCCACTTCTGCGGGGACGCTCATCGGGGA | 289 |
| 67 | ucIntPrValIleThrAlaAlaHisCysPheArgSnrThr | 84 |
| 290 | GCAGTGGGCTCGACGGCTGGCCACTGGTTCGGCAACGCTGTAGAGCT | 339 |
| 84 | erLeuTyGlnValLeuLeuGlyAlaArgGlnLeuValGlnProGlyPro | 100 |
| 340 | CCCTGTACCAAGGCTCTGCTGGGGGCAAGGACGTAAGTCAACCGGGACA | 389 |
| 101 | HisAlaMetTyrAlaArgValArgGlnValGlserrsnProIleuTyGln | 117 |
| 390 | CACGGCTATGTATGCCGGGTGAGGAGGTGAGAGCAACCCCTGTACA | 439 |
| 117 | ngIlyThrAlaSerSerAlaAspValAlaLeuValGlnLeuValaProV | 134 |
| 440 | GGGCAAGGCTCCACGCGTGAAGTGGCTGTGTGAGAGTGGAGGACACAG | 489 |
| 134 | aIProPheThrAsnTyrIleuLeuProValCysLeuProAspProSerVal | 150 |
| 490 | TGCCCTTCACCAATTACATCCCTCCCGGTGCTGCTGCACCCCTCGGTG | 539 |
| 151 | IlePheGlnThrGlyMetAsnCysTrpValIleThrGlyTrpGlySerProse | 167 |
| 540 | ATCTTTGAGACGGGATGAGTACTGCTGGGTCACTGGCTGGGGCACCCCG | 589 |
| 167 | rgIuGlnAspLeuLeuProGlnProArgIleLeuGlnIlysnLeuAlaValP | 184 |
| 590 | TGAGGAGACCTCTGCGCGAACCAGGATCTCGAGAAACTCGCTGTGTC | 639 |
| 184 | rolIleIleAspThrProIlysnCysAsnLeuLeuTyrSerIlysnAspThrGln | 200 |
| 640 | CCATATGACACACTCAAGGTGCAAACCTGCTACACCAAGCACACCGA | 689 |
| 200 | urPheGlyTyrGlnProIlysnThrIleIlysnAspPheLeuCysAlaGlyP | 217 |
| 690 | GTTTGGCTACCAACCAACCATCAAGATGACATGCTGTGCGCCGGCT | 739 |
| 217 | heGlnGlnIlysnIlysnAspAlaCysIlysnIlysnSerGlyIlysnProLeu | 233 |
| 740 | TGAGGAGGAGGCAAGAGATGCTCTCAAGGCGGACTGCGCGGCCCTTG | 789 |
| 234 | ValCysLeuValGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGln | 250 |
| 790 | GGTGTGCTCTGTGGTCACTGTGTGTGCAGGCGGGGGTGAATCACTGGGG | 839 |
| 250 | yGlnGlyCysAlaArgGlnAsnArgProGlyValTyrIleArgValThr | 267 |
| 840 | TGAGGGCTGTGCCCGCCAGAACCGGCCAGGTGTCTACTCCGTCTCAACG | 889 |
| 267 | IleHisHisAsnTrpIleHisArgIleIleProIlysnLeuGlnPheGln.Pr | 283 |
| 890 | CCACCCACAACTGGATCCATCGGAATCATCCCAACTGCAGTCCAAACCC | 939 |
| 283 | oAlaArgLeuGlyGlyGlnIlysn | 290 |
| 940 | AGCGAGGTTGGCGGCGACGAAG | 961 |
| seq_name: | /SIDS1/gcdata/geneseq/geneseqn-emb1/NA198.DAT:AAV59136 | |
| seq_documentation_block: | | |
| ID | AAV59136 standard; DNA: 980 BP. | |
| XX | AAV59136: | |
| XX | AC | |
| XX | AT | |
| XX | 07-JAN-1999 (first entry) | |
| DE | Nucleotide sequence: SF0031A, a homologue of HELA2. | |
| XX | XX | |
| XX | Serine protease; regulation; cell activity; viability; HELA2; ATC2 | |
| XX | | |

KM BCM3; testis; fertility; suppressor; testicular germ cell cancer;
 KW seminoma; testis-specific expression; antitumour; sperm development;
 XX infertility; human; chromosome 16p13.3; ss.

OS Homo sapiens.

Key Location/Qualifiers
 FH 3..899
 FT CDS /*tag= a
 FT /product= SP0031A
 FT /note= "sequence contains one internal stop codon"

XX WO9836054-A1.

PD 20-AUG-1998.

XX 13-FEB-1998; 98NO-AU00085.

XX 18-NOV-1997; 97AU-0000422.

PR 13-FEB-1997; 97AU-0005101.

AMRA-) AMRAD OPERATIONS PTY LTD.

PI Antalis TM, Hooper JD;

DR WPI; 1998-480768/41.

DR P-PSDB; AAW77304.

XX New serine protease(s) and kinase involved in regulating cell
 PT activity and viability - particularly the testis-specific protease
 PT HELA2 used for modulation of fertility and as tumour suppressor

PS Example 15; Fig 20C; 167pp; English.

XX AAV59134-36 represent HELA2 homologues. The genes are found in a cluster
 CC on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
 CC homology to serine proteases. The protein is involved in or associated
 CC with regulation of cell activity and/or viability. Administration of
 CC recombinant HELA2 (also called testisin) is used to increase fertility.
 CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
 CC testicular germ cell cancers (seminoma) and is also expressed in some
 CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is
 CC a marker/potential therapeutic target for cancer. The promoter from the
 CC HELA2 gene is useful for testis-specific expression of other genes,
 CC e.g. for gene therapy or modulation of fertility. Drugs that block
 CC activity of HELA2 should have antitumour activity (other than in
 CC testis) while in testis recombinant HELA2 should stop growth of tumours
 CC and normalise sperm development (eliminating the need for orchidectomy).
 CC Identification of mutant forms of HELA2 can be used to diagnose
 CC infertility.

Sequence 980 BP; 192 A; 325 C; 289 G; 174 T; 0 other;

alignment_scores:

Quality: 1449.00 Length: 265
 Ratio: 5.468 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-041-006-7 x AAV59136 ..

Align seg 1/1 to: AAV59136 from: 1 to: 980

26 CYSGLVATPQARMTLEUASNAETVALGIYGLINASPTHTGL 42
 3 TGTGTCGCCCCAGATCTGAACGAAATGGGGGCGACGACCA 52
 42 ngluylglutrrprrtprglnvalserileglinarqnsnglyserhisp 59
 53 GGAGGGCAGATGGCCCTGCAAGATCAGCATCAGGCGCAAGGACCACT 102
 59 hecysglyserleullealaglugintrpvalleuthralalahlis 75

103 TCGGCGGGGACAGCTCATCGCGGAGACAGTGGGTCTGACGGCTGGCGAC 152
 76 CyspharqAsnrThrSerGluThrSerLeuYrGlnValLeuLeuGlyAl 92
 153 TGTCTCCGCAACACCTGTGAGAGCTCTGTACAGGTCCTCTGGGGGC 202
 92 AARGGlnLeuValGlnProGlyProHisAlaMetYrAlaArgValArg 109
 203 AAGGACGCTAGTGTACCGCGGAGACACACCTATGATGCGCGGTGAGGC 252
 109 lvalgluserasnproleuYrGlnGlyThrAlaSerSerAlaaspAl 125
 253 AGGTGAGAGCAACCCCTGTACAGGCGACGCGCTCCAGCCTGACGTG 302
 126 AlaleuValGlnLeuGlnValAlaProValProPheThrAspYrIleu 142
 303 GCGCTGGTGAGCTGAGGACGACGAGGCTTCACCAATTAATCACTCTCC 352
 142 OVALCysLeuProAspProSerValIlePheGluThrGlyMetAsnCy 159
 353 CGTGTGCTGCTCGACCGCCCTCGGTGATCTTTGAGACGGGCGCATGACT 402
 159 rValThrGlyTrpGlySerProSerGluGluAspLeuProGluPro 175
 403 GGCTCACTGGCTGGGCGAGCCGACGTGAGGAGAGACCTCTGCCGCAAC 452
 176 ArgIleLeuGlnLeuValAlaValProIleIleAspThrProLysCys 192
 453 GGAATCTGAGAACTCGGTGCTGCTCATGACACACCAAGTGCNA 502
 192 nLeuYrSerLysAspThrGlnPheGlyYrGlnProLysThrIle 209
 503 CTGTCTACAGCAAGACACCGAGTTGGCTPCCAAACCAACCATCA 552
 209 ysasnaspmetleucysAlaGlyPheGluGlnGlyLysLysAspAlaCys 225
 553 AGATGTCATGCTGTGCGCGGCTTCAGAGAGGCGCAAGAGATGCTCG 602
 226 LysGlyAspSerGlyLysProLeuValCysLeuValGlnSerTrpLe 242
 603 AAGGGGACCTGGGGCGGCCCCCTGGTGTGCTGTGGTCACTGCGGT 652
 242 uGlnAlaGlyValIleSerTrpGlyGlnGlyCysAlaArgGlnAsnArg 259
 653 GCAGCGGGGGGTGATCAGCTGGGGGTGAGGGCTGTGCCGCCAGAACGCC 702
 259 roGlyValYrIleArgValThrAlaHisHisAsnTrpIleHisArgIle 275
 703 CAGGTGTCTATCCGTCGTACCGCCGACCACTGATCATCGGATC 752
 276 lLeuProLysLeuGlnPheGlnProAlaArgLeuGlyGlyAlnLys 290
 753 ATCCCAAACTGCAGTTCACGACGAGGAGTGGGGCGCAAGAG 797
 seq_name: /SIDSI/gcdata/geneseg/geneseg-emb1/NA2001A.DAT:AAF77000
 seq_documentation_block:
 ID AAF77000 standard; cDNA; 1130 BP.
 XX
 AC AAF77000;
 XX
 DT 29-MAY-2001 (first entry)
 DE Fusion gene of protease T in a zymogen activation vector.
 XX
 DE Human; protease T; serine protease; dermatological; desquamation;
 KW skin care; laundry; detergent; shampoo; skin flaking; fusion;
 KW zymogen activation vector; ss.
 OS Homo sapiens.
 XX Synthetic.
 XX

| | | |
|--------------------------|---|---------------------|
| 151 | IIlePheGIuThrcIyMeIaScYsTrIyValThrcIyTrpGIySerProse | 167 |
| 647 | ATTTTGGACNCGGGCATGAACTGCTGGGTACTGTGGTGGCCAGCCCCAG | 696 |
| 167 | rGIuGIuAsPLeuLeuProGIuProA'gIleUeGIuLysLeuAlaValP | 184 |
| 697 | TGAGGAAGACCTCTCGCCGACCGCGGATCGCTCAGAAATCGCTGTGC | 746 |
| 184 | rolleIleAspThrProLysCysAsnLeuTyrserLysAspThrcIu | 200 |
| 747 | CCATCATTCAGACACCCCAAGTGCACCTCTACACCAAGACACCGAG | 796 |
| 201 | PheGIuTYTrGIuProLysThrIleLysAsnAspMetLeuCysAlaGIy.. | 216 |
| 797 | TTTTGGTACCAACCCAAACCATCAAGATGACATGCTGTGGCCGCAT | 846 |
| 217 | PheGIuGIuGIyLysLysAspAlaCysLys..GIyAspSerGIyGIyProL | 233 |
| 847 | TTCGAGAGAGGGCAAGAGATGCTCGCAAGTGG.....GGAGCCA | 887 |
| 233 | eunValCysLeuValGIyGIuSerTrIleuGIuAlaGIyVal..IleSer | 248 |
| 888 | TGtGGtGGTTCATGGGGATTCAgTTCACACCGAAGGGCCAGGCCAAGG | 937 |
| 249 | TrpGIuGIuGIy.....CysAlaArgGIuAsnArgProGI | 260 |
| 938 | TGGGGTTTAGGCCCTGTTCATGGGTATGCTTCAACAGAACCGCCAG | 987 |
| 260 | yValTYrIleArgValThraIahIshIsAsnTrPIleHisArgIleIleP | 277 |
| 988 | TGCTCTTAATCCGTGTGAACGGCCCAAAACTGATCGATCGATCAATGC | 1037 |
| 277 | rolYsLeuGIuPheGIuProAlaArgLeuGIyGIyGIu | 289 |
| 1038 | CCAAACTGAG..TTCCAGCCCAACGAGTTTGGCCGCCAG | 1074 |
| seq_name: | /SID51/5cgsdata/geneseq/geneseqn_emb1/NA1999.DAT:AAx06944 | |
| seq_documentation_block: | | |
| ID | AAx06944 standard; cDNA; 1109 BP. | |
| AC | AAx06944; | |
| DT | 10-MAY-1999 (first entry) | |
| DE | Serine protease HE2NM40 homologous EST clone. | |
| XX | | |
| KW | HE2NM40; serine protease; human; cancer; inflammation; asthma | |
| KW | wasting disease; atherosclerosis; stroke; diabetes; arthritis | |
| KW | neurodegenerative disease; Alzheimer's disease; diagnosis; | |
| KW | therapy; vaccine; ss. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FT | Key | Location/Qualifiers |
| FT | CDS | 678..854 |
| XX | | /*tag= a |
| XX | | |
| XX | EP890646-72. | |
| XX | | |
| XX | 13-JAN-1999. | |
| XX | | |
| XX | 01-JUN-1998; | 98EP-0304312. |
| XX | | |
| XX | 20-FEB-1998; | 98GB-0003650. |
| XX | 10-JUN-1997; | 97GB-0012088. |
| XX | 17-OCT-1997; | 97EP-0308295. |
| XX | | |
| XX | (SMIK) SMITHKLINE BEECHAM PLC. | |
| XX | | |
| XX | Burgess NA, Clinkenbeard HE, Southan CD; | |
| XX | | |

DR MPI: 1999-072882/07.
 P-PSDB; AAM88454.
 PT New serine protease HE2NM40 polypeptides and polynucleotides -
 useful as diagnostic reagents and for prevention and treatment of
 PT cancer, Alzheimer's disease and asthma
 XX
 PS Claim 18; Page 18-19; 21pp; English.
 XX
 CC This is the nucleotide sequence of an expressed sequence tag (EST)
 CC clone. A polypeptide encoded by the clone is provided in AAM88454.
 CC Claimed polynucleotides may comprise a nucleotide sequence which
 CC has at least 70% identity to this sequence. Such polynucleotides
 CC can be obtained from a cDNA library derived from mRNA in cells of
 CC human neutrophils, adipose or synovial tissue, using expressed
 CC sequence tag analysis. A polynucleotide encoding a serine protease
 CC termed HE2NM40 (see W884453) is specifically claimed. HE2NM40
 CC polypeptides and polynucleotides are useful for diagnosing diseases
 CC related to over or underexpression of HE2NM40 protein. They can
 CC also be used to modulate HE2NM40 activity. Diseases diagnosed,
 CC prevented or treated include cancer, inflammation, asthma, wasting
 CC diseases, atherosclerosis, stroke, diabetes, arthritis, and
 CC neurodegenerative diseases and disorders including Alzheimer's
 CC disease.
 XX
 SQ Sequence 683 BP; 156 A; 205 C; 195 G; 119 T; 8 other;
 XX
 alignment_scores:
 Quality: 810.50 Length: 220
 Ratio: 4.243 Gaps: 7
 Percent Similarity: 86.818 Percent Identity: 79.091
 alignment_block:
 US-10-041-006-7 x AAX06943 ..
 Align seg 1/1 to: AAX06943 from: 1 to: 683
 78 ArgAsnThrSerGluThrSerLeu.TyrGlnValIleuLeuGlyAlaArg 94
 1 CGCAACACCTCTGAGACCTCCCTGACACAGCTCTGCTGGGGGCAAGCC 50
 94 InLeuValGlnPro.GlyProHisAlaMetYrAlaArgValArgGlnVa 110
 51 AGCTAGTCACAGCGGGGACACACGCTATGATGCGCGGAGAGGAGGT 100
 110 IGluSerAsnProLeuTyrGlnGlyThrAlaSerSerAlaAspValAlaL 127
 101 GGAGGCAACCCCTGTACAGGCGACGCGCTCCAGCGCTGACGTGGGCC 150
 127 euValGluLeuGlnAlaProValProPheThrAsnTyrIleLeuProVal 143
 151 TGGTGAGACTGGAGGACCACTGCTTACACCAATTACATCTCCCGCTG 200
 144 CysLeuProAspProSerValIlePheGluThrGlyMetAsnCytrIpa 160
 201 TGCGTGCTGACCCCTGGTATCTTTGAGAGCGGCAATGAAGCTCTGGGT 250
 160 lThGlyTrpGlySerProSerGluGluAspLeuLeuProGluProArgI 177
 251 CACTGGCGTGGGCGACCCCACTGAGGAGACCTCTGCCCAACCGCGGA 300
 177 leLeuGlnIleValAlaValProIleIleAspThrProIleCysAsnLeu 193
 301 TCTTGACAGAACTCGCTGCTGCCATCATCGACACACCAAGTGCACCTG 350
 194 LeuTyrSerIleAspThrGluPheGlyTyrGlnProIleThrIleIleYs 210
 351 CTCTACAGACAAAGACACGAGTTGCTACCAACCAAAACATCAAGAA 400
 210 nAspMetLeuCySalagly.PheGluGluGlyIleYsAspAlaCysIleYs 226
 401 TGACATGCTGTGCGCGGNTTTCGAGGAGGCGCAAGAGATGCTGCAAG 450

seq_name: /SIDS1/genedata/geneseq/geneseqn-emb1/XA2000.DAT:AAA61696
 seq_documentation_block:
 ID AAA61696 standard: cDNA; 1007 BP.
 AC
 XX AAA61696;
 XX
 DT 23-OCT-2000 (first entry)
 XX
 DE cDNA encoding human serine protease BSSP4 (hbssp4) SEQ ID NO:3.
 XX
 KW BSSP4; serine protease; human; hbssp4; mouse; mbssp4; brain;
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;
 KW oedema; dropsy; cancer; inflammation; prostate; testis; bone; ss.
 OS
 XX Homo sapiens.
 XX
 PN WO200031277-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 19-NOV-1999; 99WO-JP06472.
 XX
 PR 20-NOV-1998; 98JP-0347813.
 XX
 PA (FUSO) FUSO PHARM IND LTD.
 XX
 PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;
 XX
 DR MPI: 2000-400084/34.
 DR P-PSDB; AAB11701.
 XX
 PT Serine protease BSSP4 and antibodies recognizing BSSP4 for assay and
 PT diagnosis of diseases in which BSSP4 expression is altered
 XX
 PS Claim 4; Page 67-69; 11pp; Japanese.
 XX
 CC The invention relates to novel serine proteases designated BSSP4
 CC (AAB11700-B11709), and to nucleic acids encoding them (AAA61695-661704,
 CC AAA61799). The invention also relates to vectors and transformants
 CC comprising BSSP4 nucleic acids; transgenic animals in which the
 CC expression level of BSSP4 can be varied; and an mbssp4 knockout mouse.
 CC The invention additionally encompasses anti-BSSP4 antibodies and methods
 CC of production of such antibodies; methods of BSSP4 detection using the
 CC antibodies; and the use of BSSP4 proteins or fragments as diagnostic
 CC markers for certain medical conditions. Nucleotides encoding BSSP4 were
 CC initially isolated in a human brain cDNA library using degenerate PCR
 CC primers (AAA61714-661715) based on conserved regions of serine
 CC proteases. The BSSP4 serine proteases and nucleotides encoding them are
 CC useful in detecting homologues, mutants and polymorphic variants in
 CC biological samples (e.g., blood, urine, brain, prostate gland and testis)
 CC as diagnostic markers for diseases associated with altered BSSP4

expression levels. Such diseases include Alzheimer's disease, oedema (dropsy), cancer or inflammation of brain, prostate, testis or bone. Sequences AAA61695-A61703 and AAA61799 represent cDNAs encoding human BSSP4 variants (hBSSP4), and sequence AAA61704 represents cDNA encoding murine BSSP4 (mBSSP4).

Sequence 1007 BP; 171 A; 311 C; 320 G; 205 T; 0 other;

alignment_scores:

Quality: 739.00 Length: 285
Ratio: 3.421 Gaps: 4
Percent Similarity: 75.789 Percent Identity: 48.772

alignment_block:

US-10-041-006-7 x AAA61696 ..

align seg 1/1 to: AAA61696 from: 1 to: 1007

```

10 LeuLeuLeuLeuGys.....PheGlySerGlnArgAlaLysAl 22
   |||||||.....:|||||:|||||:|||||:|||||:|||||:
64 CTCTGCTGCTGGCGTGGACAGCCATCTCAATGCGCGCGGATACCTGT 113
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 TCCCCAGCCTGTGGGAGGCCAGCAGCGTGAACCGGGTGTGGCGGCG 163
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 LnaSPHrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 55
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 AGGACAGCAGCTGACAGCAGTGGCCCTGGATCTGAGCAGCAGAGAT 213
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 GlySerHisPheGlySerGlySerGlySerGlySerGlySerGlySer 72
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GGGACCCAGCAGTGGCGAGGCTCTCTGCTACACGCCGCGGTGATCAC 263
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 rAlaAlaHisCysPheArg...AsnThrSerGlnThrSerLeuTyGln 88
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TGGTGGCCCATGTTTCAAGAGCAACCTGACAAACCATACCTGTCTCTG 313
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 AlLeuLeuGlnAlaArgGlnLeuValGlnProGlyProHisAlaMetTy 104
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 TGCCTGGCGGAGCATTCGCCGCTGCGGAGAACCTGCTCCAGAACG 363
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 AlaArgValArgGlnValGlnSerAsnProLeuTyGln...GlyThrAl 120
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 GTGGGTTCCTGGCTGGAGAGCCCACTGTGTATCTCGAGAGGAAG 413
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 AsnSerAlaAspValAlaLeuValGlnLeuAlaProValProPheT 137
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 TGCCTGGAGACATTCGCCGCTGCGGAGCGCTGCATACAGTTCT 463
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 hIraNTyrIleLeuProValCysLeuProAspProSerValIlePheGln 153
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 CAGAGCGGCTTCGCCATCTGCTACCTGATGCTTATCCATCCACCTCT 513
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 ThrGlyMetAsnCysTrpValThrGlyTrpGlySerProSerGlnGln 170
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
514 CCAACACACCACTCTGATCTGAGCTGGGAGGAGCATCAAGATGAGT 563
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 PLeuLeuProGlnProArgIleLeuGlnLysLeuAlaValProIleLea 187
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
564 TCCCTGGCCCACTCAGACCTCGAGAGCTGAGAGCTTCTATCATTCG 613
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 sPTIrrProLysCysAsnLeuLeuTyrSerLysAspThrGlnPheGlyTyr 203
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
614 ACTGGAAGTCTGACCATCTGTACTGGCGGAGGAGCAGCAGGAGGA... 660
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 GlnProLysThrIleLysAsnAspMetLeuLysAlaGlyPheGlnGln 220
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 .....CCATCACTGAGAGCATGCTGTGTCGCGGATCTTGAGGG 701
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 YLysLysAspAlaLysGlyLysPheSerGlyGlyProLeuValCysLeu 237
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

702 GGAGCGGAGTCTTGTCTGGCGGAGCTCCGGGCGCCCTCATGTGCCAGG 751
237 aLGlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGlnGlyCys 253
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
752 TGGACGGCGCTGCTGCTGGCGGAGCATATCACTGCGGAGGAGGCTGT 801
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 AlaArgGlnAsnArgProGlyValTyrIleArgValThrAlaHisHis 270
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
802 GCGGAGCGGCACAGGCGCGGGGTCTACATCAGCCTCTGTGGCAGCGTCT 851
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 nTriPLeHisArgIleIleProLysLeuGlnPheGlnProAlaArgLeu 287
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
852 CTGGGTGAGAGATCTGTCAAGGGGTGTCAGCTCCGCGCGCGCTCAGG 901
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 YGly 288
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
902 GGGGT 906

seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA541085
seq_documentation_block:
ID   AA541085 standard; cDNA; 1352 BP.
XX
AC   AA541085;
XX
DT   17-DEC-2001 (first entry)
XX
DE   cDNA encoding novel human enzyme polypeptide #301.
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW   anti arthritic; nephrotropic; anticoagulant; SS.
XX
OS   Homo sapiens.
XX
PN   WO20015301-A2.
XX
PD   02-AUG-2001.
XX
PE   17-JAN-2001; 2001WO-US01239.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218299.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225265.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.

```

PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244675.
 PR 08-NOV-2000; 2000US-0244676.
 PR 08-NOV-2000; 2000US-0244677.
 PR 08-NOV-2000; 2000US-0244678.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0235678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 PI
 XX
 DR WPI. 2001-465566/50.
 DR P-PSDB; AAU23215.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 4; SEQ ID NO 311; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AA540785-AA541684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SO Sequence 1352 BP; 238 A; 446 C; 407 G; 261 T; 0 other;

alignment_scores:

PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235835.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249254.
 PR 17-NOV-2000; 2000US-0249255.
 PR 17-NOV-2000; 2000US-0249256.
 PR 17-NOV-2000; 2000US-0249257.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251859.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476222/51.

P-PSDB; AAU17037.

Novel polypeptides and polynucleotides useful as diagnostic reagents to
 diagnose diseases or disorders associated with aberrant expression or
 activity of polypeptides, for treating blood clotting disorder,
 haemophilia

PS Claim 1; SEQ ID No 134; 601bp; English.

CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays (ELISA). Disorders which are diagnosed or treated
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present

alignment_scores:

Quality: 739.00 Length: 285
 Ratio: 3.421 Gaps: 4
 Percent Similarity: 75.789 Percent Identity: 48.772

alignment_block:

US-10-041-006-7 x AAS26942


```

1 APPLICANT: BURGESS, NICOLA A.
2 APPLICANT: CLINKENBEARD, HELEN E.
3 APPLICANT: SOUTHAIR, CHRISTOPHER D.
4 TITLE OF INVENTION: NOVEL COMPOUNDS
5 NUMBER OF SEQUENCES: 6
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: RATNER & PRESTIA
8 STREET: P.O. BOX 980
9 CITY: VALLEY FORGE
10 STATE: PA
11 COUNTRY: USA
12 ZIP: 19482
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq for Windows Version 2.0
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/09/088,651
20 FILING DATE: JUNE 1, 1998
21 CLASSIFICATION:
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: GB9712088.5
24 FILING DATE: 10-JUNE-1997
25 APPLICATION NUMBER: EP 97308295.1
26 FILING DATE: 17-OCT-1997
27 APPLICATION NUMBER: GB 9803650.2
28 FILING DATE: 20-FEB-1998
29 ATTORNEY/AGENT INFORMATION:
30 NAME: PRESTIA, PAUL F.
31 REGISTRATION NUMBER: 23,031
32 REFERENCE/DOCKET NUMBER: GH30358
33 TELECOMMUNICATION INFORMATION:
34 TELEPHONE: 610-407-0700
35 TELEFAX: 610-407-0701
36 TELEX: 846169
37 INFORMATION FOR SEQ ID NO: 6:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 1109 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: single
42 TOPOLOGY: linear
43 MOLECULE TYPE: cDNA
44 US-09-088-651-6
45
46 alignment_scores:
47 Quality: 1169.50 Length: 295
48 Ratio: 4.515 Gaps: 5
49 Percent Similarity: 87.737 Percent Identity: 80.678
50
51 alignment_block:
52 US-10-041-006-7 x US-09-088-651-6 ..
53
54 Align seg 1/1 to: US-09-088-651-6 from: 1 to: 1109
55
56 1 MetAtgaTgProAlaIalValAProLeuLeuLeuLeuLeuCysPheGlySe 17
57 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 198 AAGAGCGCGCGCGCGCGCGGTCCCTCCTCGTGCTGTTTNNNGN 247
59
60 17 TcGlnArGAlaLysAlaIatrrAlaCysGlyArProArGmetLeuAsnA 34
61 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 248 TCAGAGGGGCCAAGACGACGACAAGCTGTGTGTCGCCCCAGATGCTGAACN 297
63
64 34 tgmEtValtGlYtGlInAsPrThrGlnGluGluGluTrpProTrpGlnVal 50
65 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 298 NAANNGTGGNNNGCGAGCACGACGAGGAGGCGAGTAGTGACCTGCAGAATC 347
67
68 51 SerIleGlnArGaNsNgLySerHisPheCysGlytSerLeuIleAlaGI 67
69 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
70 348 AGCATCCAGCGCCCAACGAAAGCCATTMTTGGCGGGGCGACTTCATCCGGCA 397
71
72 67 uGlttrrVallauThrrAlaAlaHisCysPharGArGsNrThrSergIunhrs 84

```


APPLICATION NUMBER: US/08/978,404B
 FILING DATE: 25-NOV-97
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/032,354
 FILING DATE: 04-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Plumer, Elizabeth R.
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B080L/7090
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 TELEX:
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1219 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IS-08-978-404B-7

352 G|YCysA|AAArg|pAsnAtrpProG|vva|TyrI|eArvA|Thra|ah| 3568
353 UCACAGG|AAACGG|AACCIgGCIgCAGGCAAG|GIGG|CAAGCIGGG|GAG 3570

921 GGCTGTGCTCTGCCCAACAGGCGCTGCATCTACACTCGGTCACCTATTA 970

COUNTRY: U.S.A.
 ZIP: 02210-2211
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/978,404B
 FILING DATE: 25-NOV-97
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/032,354
 FILING DATE: 04-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Plumer, Elizabeth R.
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B0801/7090
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441
 TELEX:

FORMATION FOR SEQ. ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1128 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-978-404B-15

alignment_scores:
 Quality: 588.50 Length: 274
 Ratio: 3.147 Gaps: 4
 Percent Similarity: 68.248 Percent Identity: 43.796

alignment_block:
 US-10-041-006-7 x US-08-978-404B-15 ..

Align seg 1/1 to: US-08-978-404B-15 from: 1 to: 1128

```

9  leueneuleuencysphneglyserglinarqalalysalalaltrral 25
11  CTGCTGCTGGCGCTGCCCTGCTGCGAGCGCGCTACGCGCCCTGCG 60
25  acysglyarproargmetleuasnargmetvalglyglinsphrg 42
61  CCCAGCGCCAGCCCTGCGAGGAGTGGGATCGTTGGGGTCAGAGGCC 110
42  lnglglgluttrprrprrprrprrprrprrprrprrprrprrprr 56
111  CCAGGAGCAAGTGGCCCTGCGAGTGGAGTCCAGGCGCCCATAC 160
57  ...Serishphecysglyserleuilealgluglntrprrprrprr 72
161  TGGATGACTTTCGCGGGGCTCCCTCATCCACCCAGTGGGTGCTGAC 210
72  ralaalanhiscysphe...Argasnhrserglurhrserleutyr 88
211  CGAGCGCCACTGCTGGGACCGGACGTCAGATCGGCGCCCTCAGGG 260
88  alleueneuglyalarglnleuvalglinproglprrprrprrprr 104
261  TGCAGCTGCGGAGCAGACCTCTACTAC.....CAGGACCGCTG 301
105  Alargvalarglnvalgluserrsnproleutyrleuglnlythrlase 121
302  CTCCCGGTGACGAGATCATGCGACCCACACATTTACACCGCCAGAT 351
121  rSerAlaaspvalalaleuvalgluleuglnalaprovalprophetr 138
352  CGAGCGGAGCATGCGCTGCTGAGACTGGAGGAGCGGATGAAGGTCTCA 401

```

```

138  snTyrileuProvalCysleuProaspProserValIlephleglurh 154
402  GCCACGTCCACAGGTCACCTGCCCCCTGCTCCAGACCTTCCCCCG 451
155  GlyMetAsnCystrpvalThrlyTrpGlyserProserGluAsp 171
452  GGGATCGCGCTGGGTGCTGCTGCTGCGGCGATGTCAGACATGATGAG 501
171  uLeuProgluProArgIleuGlnLysleuAlavalProIleleasPT 188
502  CCGTCCACCGCCATTCTCTGAGAGCGAGGAGGAGGAGGAGGAGGAA 551
188  hrProLysCysAsnleuLeuTyrSerLysAspThrGlnphleglur 204
552  ACCACATTTGTGACGAAATACACCTTGCGGCTACACGGGAGAGCAG 601
205  ProLysThrIleLysAsnAspmetleuCyalaGlyPheglugly 221
602  GTCCGATGCTGCGTGCAGACATGCTGTGCGCGG.....AACACCG 645
221  slrAspAlaacylsylAspserglylyProleuValCysleuValG 238
646  GAGGACTCATGCCAGCGGAGTCCGAGGCGCCCTGCTGCAAGGTGA 695
238  lylInserTrpleuglnalaglvalIleSerTrpGlyglucysla 254
696  ATGGACACTGCTGCGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCT 745
255  ArgGlnAsnArgProglValTyrIleArgValThrAlaHisAsnTr 271
746  CAGCCCAACCGCGCTGCGATTCACACCGCTGTCACCTACTACTGAG 795
271  pIleHisArgIleleProlys 278
796  GATCCACCACTATGTCCCAA 817

```

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-09-016-366A-18

seq_documentation_block:

Sequence 18, Application US/09016366A
 Patent No. 5955431
 GENERAL INFORMATION:
 APPLICANT: Stevens, Richard L.
 APPLICANT: Huang, Chifu
 TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
 TITLE OF INVENTION: INHIBITORS
 NUMBER OF SEQUENCES: 65
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
 STREET: 600 Atlantic Avenue
 CITY: Boston
 STATE: MA
 COUNTRY: U.S.A.
 ZIP: 02210-2211
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/016,366A
 FILING DATE: January 30, 1998
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/037,090
 FILING DATE: 05-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Plumer, Elizabeth R.
 REGISTRATION NUMBER: 36,637
 REFERENCE/DOCKET NUMBER: B0801/7093
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-720-3500
 TELEFAX: 617-720-2441


```

      ::  :::::|||||
600 GTCGCATGTCCTGACGACATGCTGTGCGGG...AACACCCG 643

```

Align seg 1/1 to: US-08-978-404B-13 from: 1 to: 1137

```
seq_name: /cgrn2_6/prodata/1/1na/5B_COMB.seq:US-08-978-440A-B-1
seq_documentation_block:
; Sequence 1, Application US/08978404B
; Patent No. 5968782
GENERAL INFORMATION:
```

```

1  APPLICANT: Stevens, Richard L.
2  TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
3  TITLE OF INVENTION: FIBRINOGEN
4  NUMBER OF SEQUENCES: 74
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
7  STREET: 600 Atlantic Avenue
8  CITY: Boston
9  STATE: MA
10 COUNTRY: U.S.A.
11 ZIP: 02210-2211
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Diskette
15 COMPUTER: IBM Compatible
16 OPERATING SYSTEM: DOS
17 SOFTWARE: FastSeq for Windows Version 2.0
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/978,404B
21 FILING DATE: 25-NOV-97
22 CLASSIFICATION: 435
23
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 60/032,354
26 FILING DATE: 04-DEC-1996
27
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Plummer, Elizabeth R.
30 REGISTRATION NUMBER: 36,637
31 REFERENCE/DOCKET NUMBER: B0801/7090
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 617-720-3500
34 TELEFAX: 617-720-2441
35
36 TELEX:
37
38 INFORMATION FOR SEQ ID NO: 1:
39
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 1031 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: single
44 TOPOLOGY: Linear
45
46 US-08-978-404B-1
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
```



```

1 GENERAL INFORMATION:
2 APPLICANT: Bandman, Olga
3 Hillman, Jennifer L.
4 Yue, Henry
5 Guegler, Karl J.
6 Corley, Neil C.
7 Tang, Tom Y.
8 Shah, Purvi
9
10 TITLE OF INVENTION: HUMAN PROTEASE MOLECULES
11
12 NUMBER OF SEQUENCES: 24
13
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: Incyte Pharmaceuticals, Inc.
16 STREET: 3174 Porter Dr.
17 CITY: Palo Alto
18 STATE: CA
19 COUNTRY: USA
20 ZIP: 94304
21
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Diskette
24 COMPUTER: IBM Compatible
25 OPERATING SYSTEM: DOS
26 SOFTWARE: FastSeq for Windows Version 2.0
27
28 CURRENT APPLICATION DATA:
29 APPLICATION NUMBER: US/09/008,271A
30 FILING DATE: 16-Jan-1998
31
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: <Unknown>
34 FILING DATE: <Unknown>
35
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Mohan-Peterson, Sheela
38 REGISTRATION NUMBER: 41,201
39
40 REFERENCE/DOCKET NUMBER: PF-0458 US
41
42 TELECOMMUNICATION INFORMATION:
43 TELEPHONE: 650-855-0555
44 TELEFAX: 650-845-4166
45
46 INFORMATION FOR SEQ ID NO: 15:
47 SEQUENCE CHARACTERISTICS:
48 LENGTH: 1081 base pairs
49 TYPE: nucleic acid
50 STRANDEDNESS: single
51 TOPOLOGY: linear
52
53 IMMEDIATE SOURCE:
54 LIBRARY: PROSTUT03
55 CLONE: 789927
56
57 SEQUENCE DESCRIPTION: SEQ ID NO: 15 :
58
59 US-09-008-271A-15
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
```

63 rleuilelaagluhntprvalleuthralahiscysphary ... 78
 249 COTGCTCAGCACCGCTGGGACTACACGGCGGCACCTTTGAAACT 298
 79AsnThrSerGluThrSerLeuTyrglnValIleuLeuGlyAlaArg 93
 299 ATAGGACCTTAGTGATGCCCTCGGGGTGGATGGTGCACATTGGC 342
 94 GlnLeuValGlnPro.....GlyProHisAlaMetTyrrAlaAr 106
 343 CAGCTGACTTCATCCATCCATCTTCGGAGCCGTGCAGGCTCATCACCG 392
 106 g.....ValArgGlnValGlnSerAsnProLeuArgGlnGlyThrAlas 121
 393 TTACTTCGATGGAATATCTATCTGAGCCCTCGCTACCTGGGGAAT...T 439
 121 eSerAlaAspValAlaLeuValGlnLeuGlnAlaProValProPheThr 137
 440 CACGCTATGACATTCCTCGTTGGTGAAGCTGTGCACCTGTCCACTAC 489
 138 AsnTyrrIleLeuProValCysLeuProAspProSerValIlePheGluTh 154
 490 AAACACATCCACACCCATCTGTCTCCAGGCCCTCCACATTTGATTGAA 539
 154 rGlyMetAsnCySTPrValThrGlyTyrrGlySerProSerGlnGlnAsp 171
 540 CCGGACAGACTCTCTGGTGACTGCTGGGGGTGACATCAAGAAGATGAG 589
 171 euleuProGluProArGlyIleuGlnLysLeuAlaValProIleLeasp 187
 590 CACTGCATCTCCACACACCCCTCAGGAAGTTCAAGTGCATCATTAAC 639
 188 ThrProCysAspLeuLeuTyrrSerLysAspThrGluPheGlyTyrgl 204
 640 AACTCTATGTGCAACCACTCTCTCTCAAG.....TACAGTTTCG 680
 204 nProLysThrIleLysAsnAspMetLeuCysAlaGlyPheGluGlnGly 221
 681 C...AAGGACATCTTTGGAGACATGTTGTGTGTCGCAATGACCAAGCG 727
 221 yLysAspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeuVal 237
 728 GGAAGGATCCCTGCTCGGTGACTCAGAGTGAGCCCTTGCCCTGTAAACG 777
 238 GlyGlnSerTrpLeuGlnAlaGlyValIleSerTrpGlyGluGlyCysAl 254
 778 AATGACTGTGTATCAGATTGTGAGTCGAGACTGGGGGTGGCTGTGG 827
 254 aArgGlnAsnArGProGlyValTyrrIleArgTyrrAlaHisHisAsnT 271
 828 TCGGCCCATCGGCCCGGTGTCTACCAACAATATCAGCACCACTTTGAGT 877
 271 rPleHisArgIleIle 276
 878 GGATCCAGAAAGCTGATG 894
 seq_name: /cgn2_6/ptodata/1/ina/6B_COMB.seq:US-09-079-970A-4
 seq_documentation_block:
 : Sequence 4, Application US/09079970A
 : Patent No. 6274366
 : GENERAL INFORMATION:
 : APPLICANT: Maffitt, Mark A.
 : APPLICANT: Niles, Andrew L.
 : APPLICANT: Haak-Frendscho, Mary
 : TITLE OF INVENTION: Enzymatically-Active Recombinant Human
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Intellectual Property Department
 : STREET: 8000 Excelsior Drive, Suite 401
 : CITY: Madison

```

STATE: WISCONSIN
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patenlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,970A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Leone, Joseph T.
REGISTRATION NUMBER: 37,170
REFERENCE/DOCKET NUMBER: 34506.073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 831-2100
TELEFAX: (608) 831-2106
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 771 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 7..753
NAME/KEY: misc.signal
LOCATION: 7..18
US-09-079-970A-4

```

```

alignment_scores:
Quality: 573.50      Length: 249
Ratio: 3.286         Gaps: 4
Percent Similarity: 69.880   Percent Identity: 44.578

```

alignment_block:
US-10-041-006-7 x US-09-079-970A-4

seq 1/1 to: US-09-079-970A-4 from: 1 to: 771

```

34 ArgMetValGlyGlyGlnAspThrGlnGluGlyGlnTrpProTrpGlnVal 50
16 AGAATCGTCGGGGCTGAGAGAGCCGCCAGAGCAAGTGGCCCTGGCAGGT 65
50 lSerIleGlnArgAsnGly.....SerHisPheCysGlyGlySerL 64
66 GAGCTTAGAGTCCAGCGCCCATAGTGAATGCACTTCTGGCGGGCTGCC 115
64 euIleAlaGlnGlnTrpValIleuThrAlaIleHisCysPhe...ArgAsn 79
116 TCATCCACCCCACTGAGTGTGCTGACCCAGCAGCACTCGTGGAGACGGAC 165
80 ThrSerGlnThrSerLeuTrpGlnValIleuLeuGlyAlaArgGlnLeuVal 96
166 GTCMAAGATCTGGCCCTCAGGCTGCACTGCGGAGAGCAGACCACTCTA 215
96 lGlnProGlyProHisAlaMetIleValIleArgValAlaGlnValGlnSer 113
216 CTAC.....CAGGACCAAGTCTGCTGCCGTCAGCAGGATCATCTGCTG 256
113 snProLeuTrpGlnGlyThrAlaSerSerAlaAspValAlaIleValGln 129
257 ACCCACAGTCTACACCGCCCAAGATCGAGCGAGCAATCATCGCTGCTGGAG 306

```

```

130 LeuGluAlaProValProPheThrAsnTrpIleLeuProValCysLeuPr 146
307 CTGAGAGAGCCGGTGAAGTCTTCACACGTCACACGAGTCACTGCGC 356
146 oaSPProSerValIlePheGlnTrpGlyMetAsnCysTrpValIleGlyT 163
357 CCTGCTCAGACACTTCCCGGGGAGTACCGCTGCTGGTGGTCACTGGCT 406
163 rplYSerProSerGlnGluAspLeuLeuProGlnProArgIleLeuGln 179
407 GGGGCAATGTGGACAAATGATGAGCGCTCCACCGCATTTCTCTGAAG 456
180 LysLeuAlaValProIleIleAspThrProLysCysAsnLeuLeuTyrSe 196
457 CAGTGAAGTCCCAATATGGAAGAAACCAATTTGTGAGCAAAATACCA 506
196 rlyAspThrGlnPheGlyTrpGlnProLysThrIleLysAsnAspMetL 213
507 CTTGGCGCTACACGCGAGACGAGCTCCGATCGTCCGTGACGACATGC 556
213 euCysAlaGlyPheGlnGluGlyLysLysAspAlaCysLysGlyAspSer 229
557 TGTGTCCCGGG....AACCCCGGAGGAGCTCATGCTCAGGGCGACTTCC 600
230 GlyGlyProLeuValCysLeuValGlyGlnSerTrpLeuGlnAlaGlyVa 246
601 GAGAGGCCCTGTGTGTGCAAGTGAATGCACTGCTGCTGACAGCGCGCT 650
246 lIleSerTrpGlyGlyCysAlaArgGlnAsnArgProGlyValTyrI 263
651 GTCACACTGGGGGAGGAGGCTGTGCCAGCCCAACCGGCTGCGCATCTCA 700
263 LeuValThrAlaHisAsnTrpIleHisArgIleLeuProLys 278
701 CCCGTCACTACTACTGTGAGCTGATCCACCACTATGTCCCAAA 747

```

seq_name: /cgn2_6/ptodata/1/ina/5B_COWB.seq:US-08-978-404B-4

seq documentation block:

```

Sequence 4, Application US/08978404B
Patent No. 5968782
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESS: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:

```



```

51 Ser1leGlnArgAsnGlySerHisPheCysGlyGlySerLeu1lealag1 67
|||||
205 AACATCCAGCAGCAAGCAAGCAATTCGCGGGGAGAGCTCATATGGCGGA 254
67 uGlnTrpValLeuThrAlaAlaHisCysPheArgAsnThrSerGlyThrS 84
|||||
255 GCAGGGGGTCTGAGGGGCTGCGACTGCTCCGCAACACCTCTGAGAGCT 304
84 eLeuTyr GlnValLeuLeuGlnValAlaArgGlnLeuValGlnProGlyPr 100
|||||
305 CCTGTACATGCTGCTGCTGGGGCAAGCAGCTAGTGTGCGGGAGACC 354
100 oHisAlaMetTyrAlaArgValAlaGlnValGlnSerAsnProLeuTyrG 117
|||||
355 ACACCTATGATGCTGCGGGGAGAGGAGAGGAGCAACCCCTGTAC 404
117 InGlyThrAlaSerSerAlaAspValAlaLeuValGlnLeuGlnAlaPro 133
|||||
405 AGGGAGCGGCTCCAGCGCTGACGTGGCTGTGAGCTGAGAGCACCA 454
134 ValProPheThrAsnTyrIleLeuProValCysLeuProAspProSerVa 150
|||||
455 GTGCCCTTACCAATTAATCATCTCCCTGTGCTGCTGCTGACCTCGGT 504
150 IllePheGlnThrGlyMetAsnCysTrpValThrGlyTyrPheGlySerProS 167
505 GATCTTTGAGACGGGCGATGAACTGCTGGGTCACTGGTGGGGAGACCCCA 554
167 eGlnGlnAspLeuLeuProGlnProArgGlyIleGlnGlnLeuAlaVal 183
555 GTGAGGAGAACCTCTGCGCGAAGCGGATCTGCAAAACTGCTGTG 604
184 ProIleIleAspThrProLysCysAsnLeuLeuTyrSerLysAspThrGl 200
605 CCCATCAGCAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 654
200 uPheGlyTyrGlnProLysThrIleLysAsnAspMetLeuCysAlaGly. 216
655 GTTTGGGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 704
217 PheGlnGln GlnLysLysAspAlaCysLys GlnAspSerGlyGlyPro 232
705 TTGAGGAGGGGCAAGAGATGCTGCAAGGGGAGCTGCGGGGCGCC 754
232 LeuVal CysLeu ValGlnInsTrp LeuGlnAla GlnValIleS 248
755 CTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 804
248 eTrp GlnGlnGly CysAlaArgGlnAsnArgPro GlnValTyrI 263
|||||
805 GCTGGGGGTAAAGGCTGTGCGCGCAAGAAACCGCGGAGGTGTACAA 854
263 leaArgValThrAlaHisHisAsnTrpIleHisArgIleIleProLysLeu 279
855 TCCGTGACCGCCCAACAATGATTCATCGGATCATATCCCAAAATG 904
280 GlnPheGlnProAlaArgLeuGlnGlyGlnLys 290
905 GCGGTTCACCAACCAAGAGTTGCGGCGCAAAA 938
seq_name: gb_hlc:AK014645
seq_documentation_block:
LOCUS AK014645 1323 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
library, clone:4733401N09: BRAIN SPECIFIC SERINE PROTEASE-4 TYPE1,
full insert sequence.
ACCESSION AK014645
VERSION AK014645.1 GI:12852632
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) 10 days neonate skin cDNA to mRNA,
clone: 4733401N09.

```

```

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,T.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multiplexed sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 11076861
PUBMED 120350913
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Balderelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotanli,K., Ishii,Y., Itoh,M., Izawa,M., Katsukawa,T.,
Kato,H., Kawai,T., Kojima,Y., Kono,H., Koda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,D.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in RIKEN
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGGAGAGATTCGAGTTAATTAAATTAATCCCCCCCC 3']. cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 100.0. Second strand cDNA was prepared with the primer
GAGAGGAGATTCGAGTTAATTAAATTAATTAATCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb

```


Ratio: 5.398 Gaps: 0
Percent Similarity: 96.241 Percent Identity: 95.489

alignment block:
US-10-041-006-7 x AM170323/rev ..

Align seg 1/1 to reverse of: AM170323 from: 1 to: 602

```

158 CysTrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuProGly 174
|||||
602 TGCTGGGTCTACGTCGCGGACACCTAGTGAAGTAAAGCTCTGCTGCGCA 553
174 uProArgIleLeuGlnLysLeuAlaValProIleIleAspThrProLysC 191
|||||
552 ACAGCGGATCTCTGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503
191 yAsnLeuLeuTyrrSerLysAspThrGluPheGlyTyrrGlnProLysThr 207
|||||
502 GTACACATGCTCTACACCAAGACACCGAGTTGGCTTACCAACCCAAAC 453
208 IleLysAsnAspMetLeuCysAlaGlyPheGlnGluGlyLysLysAspAl 224
|||||
452 ATCAAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403
224 acYsLysGlyAspSerGlyGlyProLeuValCysLeuValGlyGlnSerT 241
|||||
402 TTGCAAGGCGGACTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCT 353
241 rPheGlnIleGlyValIleSerTrpGlyGluGlyCysAlaArgGlnAsn 257
|||||
352 GGTGCGAGGCGGGGTGATGATGCTGGGGTGAAGGCTGTGCGCCGCAAGC 303
258 ArgProGlyValTyrrIleArgValThrAlaHisAsnTrpIleHisArg 274
|||||
302 CGCCGAGGTGATACATCGCTGTCACCGCCACCAACCACTGATCATG 253
274 gIleIleProLysLeuGlnPheGlnProAlaArgLeuGlyGlnLys 290
|||||
252 GATCATCTCCCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204
seq_name: gb_est1:AM084167

```

seq documentation block:

LOCUS AM084167 601 bp mRNA linear EST 14-OCT-1999
DEFINITION xc48c07.x1 NCI.CGAP_Eso2 Homo sapiens cDNA clone IMAGE:2587500 3'
similar to TR:043342 O43342 SP001LA ;, mRNA sequence.

ACCESSION AM084167
VERSION AM084167.1 GI:6039319
FEATURES EST.

ORIGIN

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 601)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: rgabs-remail.nih.gov
Tissue Procurement: Nan Hu, M.D., Ph.D., Mark Roth, M.D., Phillip
Taylor, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bdrrp/image/image.html

FEATURES

Possible reversed clone: polyT not found
Seq primer: -40UP from gibco
High quality sequence stop: 408.
Location/Qualifiers

```

source 1..601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2587500"
/clone_lib="NCI-CCAP_Eso2"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B"
/note="Organ: esophagus; Vector: pCMV-Sport6; Site:1: Salt
; Site:2: NCI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.1 kb. Life Technologies catalog
#: 11502-010"

BASE COUNT 114 a 157 c 203 g 127 t
ORIGIN

alignment_scores:
Quality: 667.00 Length: 137
Ratio: 5.131 Gaps: 3
Percent Similarity: 94.891 Percent Identity: 93.431

alignment_block:
US-10-041-006-7 x AM084167/rev ..

Align seg 1/1 to reverse of: AM084167 from: 1 to: 601

157 AsnCysTrpValThrGlyTrpGlySerProSerGluGluAspLeuLeuProGly 173
|||||
601 AACTGCTGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
173 rGluProArgIleLeuGlnLysLeuAlaValProIleIleAspThrPro 189
|||||
551 CGAACCGCGGATCTCGAGAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
190 Lys CysAsnLeuLeuTyrrSerLysAspThrGluPheGlyTyrrGlnPro 206
|||||
501 AAGGTGCAACCTGCTTACACCAAGACACCGAGTTGGCTTACCAACCA 452
206 yStrIleLysAsnAspMetLeuCysAlaGlyPheGlnGluGlyLysLysAspAl 222
|||||
451 AAACCATCAAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 402
223 AspAlaCysLysGlyAspSerGlyGlyProLeuValCysLeuValGlyGln 239
|||||
401 GATCGCTGCAAGGCGGACTCGGCGGCGGCGGCGGCGGCGGCGGCTGCTG 352
239 nSerTrpLeuGlnIleGlyValIleSerTrpGlyGluGlyCysAlaArgC 256
|||||
351 GTCTGCTGCTGAGGCGGGGTGATGATGCTGGGGTGAAGGCTGTGCGCCG 302
256 lnaAsnArgProGlyValTyrrIleArgValThrAlaHisAsnTrpIle 272
|||||
301 AGAACCGCCGAGGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 252
273 HisArgIleIleProLysLeuGlnPheGlnProAlaArgLeuGlyGln 289
|||||
251 CATGATCATCTCCCAACCTGCAAGTCCACCGAGGAGTTGGCGGCC 202
289 lnyLys 290
|||||
201 AGAAG 197
seq_name: gb_est1:BB115142

seq documentation block:
LOCUS BB115142 625 bp mRNA linear EST 18-OCT-2001
DEFINITION BB115142 RIKEN full-length enriched, adult male urinary bladder Mus
musculus cDNA clone 9530049D14 3', mRNA sequence.
ACCESSION BB115142
VERSION BB115142.2 GI:16261796
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Email: c9abs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM1243 row: 1 column: 14
 High quality sequence stop: 704.
 Location/Qualifiers

FEATURES
 source
 1. 889
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5101213"
 /clone_1lb="NCI_CGAP_L19"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: liver; Vector: PCMV-SPORT6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.9 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 COUNT 188 a 249 c 242 g 210 t
 GCIN

alignment_scores:
 Quality: 594.50 Length: 254
 Ratio: 3.231 Gaps: 9
 Percent Similarity: 72.441 Percent Identity: 50.394

alignment_block:
 US-10-041-006-7 x B1218460 ..

Align seg 1/1 to: B1218460 from: 1 to: 889

```

24 ThrAlaCysGlyArgProArgMetLeuAsnArgMetValGlyGlyGlnAs 40
   ::::::::::::::::::::
7 TCAGGTGTGGACGTCACGACATCCAGCATCCGTCGACGACACAGA 56
   ::::::::::::::::::::
40 PThGlnGluGluTyrProTArgLysValSerLysGlnArgAsnGly 57
   ::::::::::::::::::::
57 TGCCAGCTGGGCGGCTGGCTGGCAGGTACGCTGGGAGATGGAG 106
   ::::::::::::::::::::
57 eRhISpHeCysGlyLysSerLeuIleAlaGlnTyrValLeuThrAla 73
   ::::::::::::::::::::
107 GCACAGGTGTGGGAGATCTGTATGCTGAGGAGCTGGTGTGACGT 156
   ::::::::::::::::::::
74 AlaHisCysPheArgAsnThrSerGluThrSerLeuTyrGlnValLeu 90
   ::::::::::::::::::::
157 GCCATTGCTTCAACCATGGCCAGTCCCTGTATCAACAGTACTGCT 206
   ::::::::::::::::::::
90 uGlyAlaArgLysLeuValGlnProGly.....ProHisAlaMetTyr 105
   ::::::::::::::::::::
207 GGGTACCATCTCTCTACCTGAGGACATGACCCAGAGAGCTGGAG 256
   ::::::::::::::::::::
105 laarGValArgGlnValGlnSerAsnProLeuTyrGlnGlyThrAla... 120
   ::::::::::::::::::::
257 CT...GTGGCCAGCTTATCAAGCACCAGCTATTCAGCGAGAGACAC 303
   ::::::::::::::::::::
121 SerSerAlaAspValAlaLeuValGluLeuGlnAlaProValProPhe 137
   ::::::::::::::::::::
304 AGCGATGGAGCATTTGCCCTGTGACCTGCTTACCCATCTCTCTCAA 353
   ::::::::::::::::::::
137 rAsHTyrIleLeuProValCysLeuProAspProSerValIlePheGlu 154
   ::::::::::::::::::::
354 TGACTTACATGCTTCCAGTCTGCTCCGAAACCTGGGAGACCCCTGATC 403
   ::::::::::::::::::::
154 hrGlyMetAsnCysTyrValThrGlyTyrPheSerProSerGluGluAsp 170
   ::::::::::::::::::::
404 CTGGCACCATTGCTGCTGGGTCACTGGGAGACATTTGGCACAAATCAA 453
   ::::::::::::::::::::
171 LeuLeuProGluProArgIleLeuGlnLysLeuAlaValProIleIleAs 187
  
```

```

454 CCCTCCACACACCTTTACCTGCGAGATTCGACGTGCTCTCATTTCA 503
   ::::::::::::::::::::
187 pRhProLysCysAsnLeuTyrSerLysAspRhGlnuphGlyTyrG 204
   ::::::::::::::::::::
504 TGCCGAGACCTGCATTAATCAAGGAGACTCCATTCCTGGCAGCG 553
   ::::::::::::::::::::
204 InProLysThrIleLysAsnAspMetLeuCysAlaIlePheGluGlu 220
   ::::::::::::::::::::
554 AGCA...GTATCTCTTGAAGCATGCTGTGCTGTTCCAGGAAGGC 600
   ::::::::::::::::::::
221 LysAlaAspAlaCysLys...GlyAspSerGlyGlyProLeuValCys 237
   ::::::::::::::::::::
601 AAGAGATCTGTTGAATGCTGATCCGAGAGCTCCCTAGTCTGACAA 650
   ::::::::::::::::::::
237 aLgIleGlnSer...TyrLeu...GlnAlaGlyValIleSerTyrPhe 651
   ::::::::::::::::::::
651 TTCATGACTGCTGTGATCCAGCAGGAGGTGCTGATGGGATCTGTAT 700
   ::::::::::::::::::::
252 LysAlaAlaArgLysAsnArgPro.....GlyVal 261
   ::::::::::::::::::::
701 TGTGATCTGTTCCAGAGGCGCACGCTTTACACCCAGGCTGATGTC 750
   ::::::::::::::::::::
262 TyrIle 263
   ::::::::::::::::::::
751 TACATT 756
  
```

seq_name: gb_est2:BG962187

seq_documentation_block:

LOCUS BG962187 802 bp mRNA linear EST 12-JUN-2001
 DEFINITION 602826954F1 NCI_CGAP_C024 Mus musculus cDNA clone IMAGE:4981789 5',
 mRNA sequence.
 ACCESSION BG962187
 VERSION BG962187.1 GI:14349824

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

FEATURES

source

1. 802
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4981789"
 /clone_1lb="NCI_CGAP_C024"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: colon; Vector: PCMV-SPORT6; Site.1: NotI;
 Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.6 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 179 a 213 c 234 g 176 t
 ORIGIN

alignment_scores:
 Quality: 591.50 Length: 248

204 InProlyThrIleLysAsnAspMetLeuCysAlaGlyPheGluGlu 220
 612 ACCCA...GTCATCCTTGAAGCATGCTGTGCTGTTCCAGGAAGC 658
 221 LysLysAspAlaCysGlySerGlyGlyProLeuValCysLeuV 237
 659 AAGAGGATGCTTGAATGCTATTCGGAGGCTCCCTAGTCTGTGGC 708
 237 aGlyGlnSerTrpLeuGln 243
 709 ATTATGATCTGTGATCCAG 728

seq_name: gb_est1:AL578261

seq_documentation_block:

LOCUS AL578261 936 bp mRNA linear EST 16-FEB-2001
 DEFINITION AL578261 LTI_NFL006_PL2 Homo sapiens cDNA clone CSODK002YM07 3

PRIME, mRNA sequence.

AL578261

AL578261.1 GI:12942167

EST.

DRDS

CE

ORGANISM

human.

REFERENCE

1 (bases 1 to 936)

Li, M. B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

Location/Qualifiers

1.936

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CSODK002YM07"

/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and

cloned into the Not I and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact: Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax: (1) 301 610 8371

Email: fliang@life.com URL:

http://fulllength.invitrogen.com

BASE COUNT 206 a 256 c 288 g 180 t 6 others

ORIGIN

alignment_scores:

Quality: 567.00 Length: 268

Ratio: 3.048 Gaps: 7

Percent Similarity: 69.403 Percent Identity: 43.657

alignment_block:

US-10-041-006-7 x AL578261/rev ..

Align seg 1/1 to reverse of: AL578261 from: 1 to: 936

18 GlnAgaAlaLysAlaAlaThrAlaCysGlyArgProAlaGlyMetLeuAsnArg 34

931 GAGGGGCGCGCTTATCAGACGACGCGCGGCGGCTGACATCCTCCG 882

34 GmetValGlyGlnAspThrGlnGluGlyGlnTrpProTrpGlnValS 51

861 CATCTGGGTGGAGAGGACGCCGAACTCGGGCGCTTGGCGGAGGGA 832

51 erIleGlnArgAsnGlySerHisPheCysGlyGlySerLeuIleAlaGlu 67
 831 GCCTGGCCCTGTGGAGATTCCACGTATGCGGAGGAGACCTGTCCAGCAG 782
 68 GlnTrpValLeuThrAlaAlaHisCysPheArg.....AsnThrSe 81
 781 CGCGGGAGACTCAGCGGGGCGGACCTGTTGAAACCTATGATGACCTTAG 732
 81 rGluThrSerLeuTrpGlnValLeuGlnAlaArgGlnLeuValGlnP 96
 731 TGATCCCTCCGGGTGATGGTCCAGTTGGC.....CAGCTACTTCCA 688
 98 r.....GlyProHisAlaMetTrpAlaArg.....ValArg 108
 687 TGCCATCCTTCTGAGAGCTGAGAGCTTACACACCTGTTMTTCGTATGC 638
 109 GlnValGlnSerAsnProLeuTrpGlnGlyTrpAlaSerSerAlaSpA 125
 637 AATATCTATCTGAGCCCTCGCTACCTGGGGAAT...TCACCCATGACAT 591
 125 LAlaLeuValGlnLeuGlnAlaProValProPheThrAsnTrpIleuP 142
 590 TGCCCTGTGAGAGCTGTCTGACCTGTGACCTTACCTAAACATATCCAGC 541
 142 roValCysLeuProAspProSerValIlePheGlnTrpGlyMetAsnCys 158
 540 CCATCTGTCTCCAGGCTCCACATTGAGTTGAGAAACCGGACCAACTGC 491
 159 TrpValThrGlyTrpGlySerProSerGlnGlnAspLeuProGluTr 175
 490 TGGGTACTGCTGGGGGTACATCAAGAGATGAGGCACTGCAATCTCC 441
 175 oArgIleGlnGlnLeuAlaValProIleIleAspThrProGlyCysA 192
 440 CCAACCTCTCAGGAAGTTCAAGTCCGCATCATTAACAACTATATGCA 391
 192 snLeuLeuTrpSerLysAspThrGlnPheGlyTrpGlnProLysThrIle 208
 390 ACCACACTCTCCCTCAG.....TACAGTTTCCGC...AAGGACATC 353
 209 LysAsnAspMetLeuCysAlaGlyPheGluGlyLysLysAspAlaC 225
 352 TTGGAGACATGCTGTGCTGCAATGCCCAAGGAGGAGATGCTG 303
 225 sLysGlyAspSerGlyGlyProLeuValCysLeuValGlnSerTrpL 242
 302 CTTCGTAATCACTGACGTGACCTTGGCTGTACAGAAAGACGCTGTGT 253
 242 euGlnAlaGlyValIleSerTrpGlyGlnGlyCysAlaArgGlnAsnArg 258
 252 ATCAGATTGAGTGTGACGCTGGGAGTGTGTGCTGCGCCCAATCG 203
 259 ProGlyValTrpIleArgValThrAlaHisAsnTrpIleHisArgI 275
 202 CCGGTGTCTACACCAATATACCAACCACTTTGAGTGATCCAGAACT 153
 275 eIle 276
 152 GATG 149

seq_name: gb_hc:AK006271

seq_documentation_block:

LOCUS AK006271 1050 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:1700023E12;protease, serine, 21, full insert

sequence.

ACCESSION AK006271

VERSION AK006271.1 GI:12839279

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,

clone:1700023E12.

| | |
|--|---|
| | ORGANISM |
| Mus musculus | |
| Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| REFERENCE | |
| AUTHORS | 1 (sites) |
| JOURNAL | Carrincci,P. and Hayashizaki,Y. |
| PUBMED | High-efficiency full-length cDNA cloning |
| REFERENCE | Meth. Enzymol. 303, 19-44 (1999) |
| AUTHORS | 99279253 |
| TITLE | 10349636 |
| JOURNAL | 2 (sites) |
| PUBMED | Carrincci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., |
| REFERENCE | Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y. |
| AUTHORS | Normalization and subtraction of cap-trapper-selected cDNAs to |
| TITLE | prepare full-length cDNA libraries for rapid discovery of new genes |
| JOURNAL | Genome Res. 10 (10), 1617-1630 (2000) |
| PUBMED | 20499374 |
| REFERENCE | 11042159 |
| AUTHORS | 3 (sites) |
| TITLE | Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carrincci,P., |
| JOURNAL | Kono,H., Akizawa,T., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., |
| PUBMED | Sunil,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., |
| REFERENCE | Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasbiwagi,K., |
| AUTHORS | Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Onara,E., Mchikiri,M., |
| TITLE | Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,K., Tanaka,T., Matsura,S., Kawaj,J., |
| JOURNAL | Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. |
| PUBMED | Riken integrated sequence analysis (RISA) system--384-format |
| REFERENCE | sequencing pipeline with 384 multicapillary sequencer |
| AUTHORS | Genome Res. 10 (11), 1757-1771 (2000) |
| TITLE | 11076861 |
| JOURNAL | 4 (sites) |
| PUBMED | The RIKEN Genome Exploration Research Group Phase II Team and the |
| REFERENCE | FANTOM Consortium. |
| AUTHORS | Functional annotation of a full-length mouse cDNA collection |
| TITLE | Nature 409, 685-690 (2001) |
| JOURNAL | 5 (bases 1 to 1050) |
| PUBMED | Adachi,U., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., |
| REFERENCE | Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., |
| AUTHORS | Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanganai,T., |
| TITLE | Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiroaka,T., Horii,F., |
| JOURNAL | Hume,D., Imochi,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., |
| PUBMED | Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koza,S., |
| REFERENCE | Kutirera,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., |
| AUTHORS | Nunahazi,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., |
| TITLE | Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., |
| JOURNAL | Schrimal,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., |
| PUBMED | Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., |
| REFERENCE | Tanaka,T., Tejima,Y., Toyota,T., Yamanura,T., Yamanaka,I., |
| AUTHORS | Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and |
| TITLE | Hayashizaki,Y. |
| JOURNAL | Direct Submission |
| PUBMED | Submitted (10-JUL-2000) yoshihide Hayashizaki, The Institute of |
| REFERENCE | Physical and Chemical Research (RIKEN), Laboratory for Genome |
| AUTHORS | Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), |
| TITLE | RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, |
| JOURNAL | Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, |
| PUBMED | URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, |
| REFERENCE | Fax:81-45-503-9216) |
| AUTHORS | Please visit our web site (http://genome.gsc.riken.go.jp/) for |
| TITLE | further details. |
| JOURNAL | cDNA library was prepared and sequenced in Mouse Genome |
| PUBMED | Encyclopedia Project of Genome Exploration Research Group in Riken |
| REFERENCE | Genomic Sciences Center and Genome Science Laboratory in Riken. |
| AUTHORS | Division of Experimental Animal Research in Riken contributed to |
| TITLE | prepare mouse tissues. First strand cDNA was primed with a primer |
| JOURNAL | [5' GAGAGACAAAGATCCAGAAGCTCTTTTTCCTTTTTTTTNN 3'], cDNA was |
| PUBMED | prepared by using trehalose activated reverse transcriptase |
| REFERENCE | and subsequently enriched for full-length by cap-trapper. Second |
| AUTHORS | strand cDNA was prepared with the primer adapter of sequences[5' |
| TITLE | GAGAGACAGCGGCCCAATTATTCGCAGTAATTAATTAATTCATCCCCCCC 3']. cDNA |
| JOURNAL | was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI, 3' |
| PUBMED | end: SstI. Host: SOLR. |
| REFERENCE | Location/Qualifiers |

```

source
    1..1050
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="MGD:MGI:1901580"
        /db_xref="taxon:10090"
        /clone="1700023E12"
        /sex="male"
        /tissue.type="testis"
        /clone.lib="RIKEN full-length enriched mouse cDNA library"
        /dev_stage="adult"
        .943
        /gene="Prss21"
        .943
        /gene="Prss21"
        /note="data source: MGD, source key: MGI:191698, evidence: ISS protease, serine, 21 putative"
        /codon_start=1
        /protein_id="BAB24495.1"
        /db_xref="GI:12839280"
        /translation="MALOSTYLOVDPEKPELQEDLLSGPCGHRTPSRIVGGDDAEAFEGWPMQSILRWGMHGLIGARTLNRRWLTAHQFQDNQPPMTVOFGELTSRSLAMLDASINRKIEDIFLSPRYSQQNDIALKLSSPTYNNEFTOPICLANSTYKKPNRRDCMVGWGAIGEDESLPSPNTLQEVVALIINSNCMHMKKDFRTNIMGDVCAGTIEGKNACFDSDGGPLACDDPDTVYQGVVSMWIGCGRHNRPGVTYNIISHYNMIQSTIRNGLRDPVPYLFLLEFLTIA"
        .1037
        polyA_signal
            /note="putative"
        polyA_site
            /note="putative"
BASE COUNT      230 a      291 c      279 g      250 t
ORIGIN

alignment_scores:
    Quality: 560.50          Length: 312
    Ratio:   2.935           Gaps:     8
Percent Similarity: 61.218    Percent Identity: 38.141

alignment_block:
US-10-041-006-7 x AK006271 ..

Align seg 1/1 to: AK006271 from: 1 to: 1050

2 ArgatgProalaalavalProleuleuleuleucysPhneglysergl 18
||| ::::::::::::::::::::
5 CGGGGCAGACGCTTGTCGCCACTGCTGTTGTC..... 40
18 nArgAlaAlaAlaThrAla..... 25
||| ||||| |
41 ... GCGACTCGGGCATGGCCATTACAGTCACCTATTTGCAGGTGATC 86
26 ..... Cysgly 27
||||| 
87 CTGGAATAACCGAACTCGAGAACCOCGACCTATTGTCAGGGCCCTGCGGT 136
28 ArGPProArMetLeuasnArqMetValIgLyGlInAsPTHrGlInGl 44
::: :::::::::::::::::::
137 CACGAGAACCATCCCTCCCGATAGTGGTGGCGAATGATGTCGAGCTTGG 186
||||| |
44 yGUUrPrPrTrpGlnValSerilegInarGaSnGlySerHisPhncYas 61
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
187 CCCTGCCCCGTCGCAAGGAGACCTGCCTGATATGGGGCACMACCATTTATG 236
61 lyGISeLeuIlleaIgluInlrryValleuthralaalahitscysphe 77
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
237 GGCAACCTTGCTCAACCGCGCTGGTGCTTACAGCTGCCACACTGCTTC 286
78 Argasn.....Th 80
:::~
287 CAAGAAGATTAACGATCTTTTGACTGCAGATCCACTTTGGTAGCTGAC 336

```

| ACCESSION VERSION KEYWORDS SOURCE | ORGANISM | REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED REFERENCE AUTHORS |
|---|---|---|
| AK004939 AK004939.1 HTC, CAP trapper. Mus musculus (Strain: C57BL/6J) adult male liver cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library clone:1300008a22. | Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | 1 (sites) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) |
| 10349636 2 (sites) | | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. |

| | |
|-----------|---|
| TITLE | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes |
| JOURNAL | Genome Res. 10 (10), 1617-1630 (2000) |
| PUBLISHED | 20493574 11042159 |
| REFERENCE | 3 (sites) |
| AUTHORS | Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,Y., Nishi,K., Kitsuma,T., Tashiro,H., Itoh,A., Sumi.N., Ishii.Y., Nakamura.S., Hazama.M., Nishino.T., Harada.A., Yamamoto.R., Matsumoto.H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue.K., Togawa.Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda.Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuyama.S., Kawai,J., Okazaki.Y., Muramatsu.M., Inoue.Y., Kita,A. and Hayashizaki.Y. |
| TITLE | Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer |
| JOURNAL | Genome Res. 10 (11), 1757-1771 (2000) |
| PUBLISHED | 20530913 11076861 |
| REFERENCE | 4 (sites) |
| AUTHORS | The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. |
| TITLE | Functional annotation of a full-length mouse cDNA collection |
| JOURNAL | Nature 409, 685-690 (2001) |
| PUBLISHED | 5 (bases 1 to 3030) |
| REFERENCE | 5 (bases 1 to 3030) |
| AUTHORS | Adachi,U., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Komura,K., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numataki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schraml,L., Shibata,K., Shibata,Y., Shnegawa,A., Shiraki,T., Sobgar,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. |
| TITLE | Direct Submission |
| JOURNAL | Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) |
| COMMENT | Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGGAGAGACCGCCGCCAATCTGGATTGTATTTTATTTVN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of Sequence15' GAGGAGAGAAGACTCAAGAGCTCAATTAATTTATTTAAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOBR. |
| FEATURES | Location/Qualifiers 1..3030 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="MGD:MGI:1904397" /db_xref="taxon:10090" /clone="1300008A22" /sex="male" /tissue_type="Liver" /clone_lib="RIKEN Full-length enriched mouse cDNA library" /dev_stage="adult" 91..2490 /note="data source:SPRR, source key:095519, evidence:ISS |
| CDS | |

Wed Aug 14 08:39:18 2002

us-10-041-006-7.p2n.rst

Page 16